

**PENANGANAN DATA TIDAK SEIMBANG DENGAN ALGORITMA
SMOTE UNTUK KLASIFIKASI PENGANGGURAN DI
KABUPATEN LIMA PULUH KOTA MENGGUNAKAN
METODE CART**



**ALDWI RIANDHOKO
NIM. 20337002**

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2024**

**PENANGANAN DATA TIDAK SEIMBANG DENGAN ALGORITMA
SMOTE UNTUK KLASIFIKASI PENGANGGURAN DI
KABUPATEN LIMA PULUH KOTA MENGGUNAKAN
METODE CART**

SKRIPSI

*Diajukan sebagai salah satu persyaratan guna memperoleh gelar
Sarjana Statistika*



Oleh:

ALDWI RIANDHOKO

NIM. 20337002

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2024**

PERSETUJUAN SKRIPSI

PENANGANAN DATA TIDAK SEIMBANG DENGAN ALGORITMA *SMOTE* UNTUK KLASIFIKASI PENGANGGURAN DI KABUPATEN LIMA PULUH KOTA MENGGUNAKAN METODE *CART*

Nama : Aldwi Riandhoko
NIM : 20337002
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 08 November 2024

Mengetahui:
Kepala Departemen Statistika



Dr. Yenni Kurniawati, S.Si., M.Si.
NIP. 198402232010122005

Disetujui Oleh:
Pembimbing



Dra. Nonong Amalita, M.Si.
NIP. 196906151993032001

PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Aldwi Riandhoko
NIM : 20337002
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

PENANGANAN DATA TIDAK SEIMBANG DENGAN ALGORITMA *SMOTE* UNTUK KLASIFIKASI PENGANGGURAN DI KABUPATEN LIMA PULUH KOTA MENGGUNAKAN METODE *CART*

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

Padang, 08 November 2024

Tim Penguji

Nama

Ketua : Dra. Nonong Amalita, M.Si.

Anggota : Dodi Vionanda, M.Si., Ph.D.

Anggota : Admi Salma, S.Pd., M.Si.

Tanda Tangan



SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan dibawah ini :

Nama : Aldwi Riandhoko
NIM : 20337002
Program Studi : SI Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan, bahwa skripsi saya dengan judul "Penanganan Data Tidak Seimbang dengan Algoritma *SMOTE* untuk Klasifikasi Pengangguran di Kabupaten Lima Puluh Kota Menggunakan Metode *CART*" adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dan 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 masyarakat dan negara. Demikian pernyataan ini saya buat dengan penuh kesadaran dan rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Padang, 08 November 2024

Diketahui Oleh:
Kepala Departemen Statistika



Dr. Yenni Kurniawati, M.Si.
NIP. 198402232010122005

Saya yang menyatakan,



Aldwi Riandhoko
NIM. 20337002

Penanganan Data Tidak Seimbang dengan Algoritma *SMOTE* untuk Klasifikasi Pengangguran di Kabupaten Lima Puluh Kota Menggunakan Metode *CART*

Aldwi Riandhoko

ABSTRAK

Salah satu daerah yang masih mengalami masalah pengangguran di Sumatera Barat adalah Kabupaten Lima Puluh Kota. Pasca pandemi, pengangguran di Kabupaten Lima Puluh Kota mengalami kenaikan disetiap tahunnya. Pengangguran di Kabupaten Lima Puluh Kota disebabkan oleh rendahnya kompetensi sumber daya manusia untuk memenuhi kebutuhan lapangan pekerjaan. Kabupaten Lima Puluh Kota memiliki jumlah angkatan kerja yang bukan pengangguran lebih banyak dibandingkan dengan angkatan kerja yang pengangguran, sehingga mengakibatkan data yang tidak seimbang. Tujuan penelitian ini yaitu untuk mengetahui hasil pohon klasifikasi, karakteristik, dan ketetapan hasil klasifikasi dari pengangguran di Kabupaten Lima Puluh Kota sebelum dan sesudah diterapkan *SMOTE*.

Metode yang dapat mengatasi data yang tidak seimbang adalah teknik *SMOTE*. Kondisi ketidakseimbangan data perlu ditangani agar dapat meningkatkan performa model klasifikasi. *CART* merupakan salah satu teknik klasifikasi dengan metode pohon keputusan yang dapat memperoleh karakteristik dari suatu klasifikasi. Jenis penelitian yang digunakan adalah penelitian terapan dengan menggunakan data sekunder. Data yang digunakan merupakan data pengangguran di Kabupaten Lima Puluh Kota yang diperoleh melalui hasil Sakernas pada bulan agustus 2023.

Berdasarkan analisis yang telah dilakukan, karakteristik pengangguran yang diperoleh dari metode *CART* sebelum diterapkan *SMOTE* yaitu 3 untuk kategori bukan pengangguran dan 1 untuk kategori pengangguran. Sedangkan sesudah diterapkan *SMOTE* yaitu 3 untuk kategori bukan pengangguran dan 3 untuk kategori pengangguran. Penerapan metode *SMOTE* pada klasifikasi *CART* dapat meningkatkan nilai akurasi pada kelas minoritas sebesar 33% sehingga menaikkan nilai AUC sebesar 7%. Maka dapat dikatakan bahwa metode *CART* sesudah diterapkan *SMOTE* memberikan kinerja yang lebih baik sehingga metode *CART* sesudah diterapkan *SMOTE* merupakan metode terbaik dalam mengklasifikasikan pengangguran di Kabupaten Lima Puluh Kota.

Kata Kunci: *CART*, Kabupaten Lima Puluh Kota, Pengangguran, *SMOTE*

Handling Unbalanced Data with SMOTE Algorithm for Unemployment Classification in Lima Puluh Kota Regency Using CART Method

Aldwi Riandhoko

ABSTRACT

One of the regions that is still experiencing unemployment problems in West Sumatera is Lima Puluh Kota Regency. After the pandemic, unemployment in Lima Puluh Kota Regency has increased every year. Unemployment in Lima Puluh Kota Regency is caused by the low competence of human resources to meet the needs of employment. Lima Puluh Kota Regency has more unemployed labor force than employed labor force, so this results in unbalanced data. The purpose of this study is to determine the results of the classification tree, characteristics, and the accuracy of the classification results of unemployment in Lima Puluh Kota Regency before and after applying SMOTE.

The method that can address unbalanced data is the SMOTE technique. Data imbalance conditions need to be addressed in order to improve the performance of the classification model. CART is one of the classification techniques with the decision tree method that can obtain the characteristics of a classification. The type of research used is applied research using secondary data. The data used is unemployment data in Lima Puluh Kota Regency obtained through Sakernas results in August 2023.

Based on the analysis that has been done, the unemployment characteristics obtained from the CART method before applying SMOTE are 3 for the non-unemployment category and 1 for the unemployment category. Meanwhile, the characteristics after applying SMOTE are 3 for the non-unemployment category and 3 for the unemployment category. Application of the SMOTE method to CART classification can increase the accuracy value in the minority class by 33%, thus increasing the AUC value by 7%. Therefore, it can be concluded that the CART method after SMOTE is applied provides better performance so that the CART method after SMOTE is applied is the best method in classifying unemployment in Lima Puluh Kota Regency.

Keywords: CART, Lima Puluh Kota Regency, SMOTE, Unemployment

KATA PENGANTAR

Bismillahirrahmanirrahiim, Alhamdulillahirrabil'alamiin, segala puji dan syukur penulis ucapkan kepada Allah SWT yang telah memberikan rahmat dan karunia-Nya kepada penulis sehingga penulis dapat menyelesaikan penyusunan Skripsi yang berjudul **“Penanganan Data Tidak Seimbang dengan Algoritma SMOTE untuk Klasifikasi Pengangguran di Kabupaten Lima Puluh Kota Menggunakan Metode CART”** dapat terselesaikan dengan baik. Selanjutnya, shalawat beserta salam untuk nabi Muhammad SAW sebagai suri tauladan bagi seluruh umat.

Skripsi ini disusun untuk memenuhi salah satu persyaratan memperoleh gelar Sarjana pada Program Studi S1 Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang. Dalam penulisan Skripsi ini, penulis banyak mengalami kendala, namun berkat bantuan, bimbingan, kerjasama dari berbagai pihak yang bersifat membangun, dan berkah dari Allah subhanahu wa ta'ala sehingga kendala-kendala yang dihadapi dapat diatasi. Dalam kesempatan ini, penulis mengucapkan terimakasih kepada:

1. Ibu Dra. Nonong Amalita, M.Si. Selaku Pembimbing Akademik dan Pembimbing Skripsi yang telah memberikan bimbingan dan dorongan selama masa kuliah dan telah berkenan meluangkan waktu, ilmu, dan tenaga, dalam memberikan arahan selama penyusunan skripsi.
2. Bapak Dodi Vionanda, M.Si., Ph.D. sebagai Dosen Penguji Skripsi yang telah memberikan arahan dan masukan selama penyusunan skripsi.

3. Ibu Admi Salma, S.Pd., M.Si. sebagai Dosen Penguji Skripsi yang telah memberikan arahan dan nasehat selama penyusunan skripsi.
4. Bapak dan Ibu Dosen, Staf Pengajar Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
5. Kedua orang tua, abang dan adik yang selalu mendoakan, memberi semangat, dan nasihat dalam menyelesaikan skripsi ini.
6. Serta semua sahabat dan rekan-rekan Statistika 2020 yang telah banyak membantu dan memberi dukungan kepada penulis.

Semoga segala bimbingan, bantuan, dan motivasi yang telah diberikan menjadi amal kebaikan dan mendapatkan balasan dari Allah subhanahu wa ta'ala. Penulis telah berusaha semaksimal mungkin untuk memberikan yang terbaik dalam penulisan Skripsi ini, namun penulis mengharapkan saran dan kritik yang membangun demi kesempurnaan Skripsi ini. Sehingga, penulis berharap agar Skripsi ini dapat bermanfaat bagi semua pihak.

Padang, November 2024

Penulis

Aldwi Riandhoko

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR.....	iii
DAFTAR ISI.....	v
DAFTAR TABEL	vii
DAFTAR GAMBAR.....	viii
DAFTAR LAMPIRAN	ix
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Batasan Masalah.....	7
C. Rumusan Masalah	8
D. Tujuan Penelitian	8
E. Manfaat Penelitian	9
BAB II KERANGKA TEORITIS.....	10
A. <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	10
B. <i>Classification and Regression Tree (CART)</i>	11
C. Tabel Ketetapan Hasil	17
D. <i>Receiver Operating Characteristic (ROC)</i>	19
E. Pengangguran.....	20
F. Faktor-faktor yang Mempengaruhi Pengangguran	20
BAB III METODOLOGI PENELITIAN	27
A. Jenis Penelitian.....	27
B. Jenis dan Sumber Data	27
C. Struktur Data	27
D. Variabel Penelitian	28
E. Teknik Analisis	29
BAB IV HASIL DAN PEMBAHASAN.....	32
A. Deskripsi Data.....	32
B. Klasifikasi Pengangguran di Kabupaten Lima Puluh Kota.....	38
C. Pembahasan.....	51

BAB V KESIMPULAN.....	55
A. Kesimpulan	55
B. Saran.....	57
DAFTAR PUSTAKA.....	58

DAFTAR TABEL

Tabel	Halaman
1. <i>Confusion Matrix</i>	18
2. Struktur Data	27
3. Variabel Penelitian	28
4. Jumlah Angkatan Kerja pada Setiap Kategori Variabel Usia	39
5. Jarak Antar Kategori pada Setiap Variabel	40
6. Calon Simpul Kiri dan Kanan	41
7. Jumlah Angkatan Kerja Pengangguran berdasarkan Simpul 1	41
8. Hasil Perhitungan Nilai Probabilitas pada Semua Simpul	42
9. Hasil Perhitungan <i>Indeks Gini</i> pada Semua Calon Simpul	43
10. Hasil Perhitungan <i>Goodness of Split</i>	44
11. Karakteristik Pengangguran pada Pohon Klasifikasi <i>CART</i> sebelum Diterapkan <i>SMOTE</i>	45
12. <i>Confusion Matrix</i> dari Metode <i>CART</i> sebelum Diterapkan <i>SMOTE</i>	46
13. Karakteristik Pengangguran pada Pohon Klasifikasi <i>CART</i> sesudah Diterapkan <i>SMOTE</i>	48
14. <i>Confusion Matrix</i> dari Metode <i>CART</i> sesudah Diterapkan <i>SMOTE</i>	49
15. Perbandingan Ketetapan Klasifikasi <i>CART</i> sebelum dan sesudah Diterapkan <i>SMOTE</i>	50

DAFTAR GAMBAR

Gambar	Halaman
1. Tingkat Pengangguran di Kabupaten Lima Puluh Kota pada Tahun 2019-2023	3
2. Persentase Pengangguran dan Bukan Pengangguran di Kabupaten Lima Puluh Kota Tahun 2023	4
3. Pohon Klasifikasi <i>CART</i>	13
4. Diagram Alir Penelitian	31
5. Deskripsi Data Pengangguran di Kabupaten Lima Puluh Kota	32
6. Mosaik Plot dari Variabel Usia dan Pengangguran	33
7. Mosaik Plot dari Variabel Jenis Kelamin dan Pengangguran	34
8. Mosaik Plot dari Variabel Status Perkawinan dan Pengangguran	34
9. Mosaik Plot dari Variabel Status dalam Keluarga dan Pengangguran	35
10. Mosaik Plot dari Variabel Tingkat Pendidikan dan Pengangguran	36
11. Mosaik Plot dari Variabel Pengalaman Kerja dan Pengangguran	37
12. Mosaik Plot dari Variabel Daerah Tempat Tinggal dan Pengangguran	38
13. Pohon Klasifikasi <i>CART</i> sebelum Diterapkan <i>SMOTE</i>	45
14. Data <i>Training</i> yang Telah Diseimbangkan	47
15. Pohon Klasifikasi <i>CART</i> sesudah Diterapkan <i>SMOTE</i>	48
16. Pohon Klasifikasi <i>CART</i> Sebelum dan Sesudah Diterapkan <i>SMOTE</i>	52

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Penelitian	61
2. Codingan <i>CART</i> sebelum diterapkan <i>SMOTE</i>	63
3. Codingan <i>CART</i> sesudah diterapkan <i>SMOTE</i>	64
4. Perhitungan Jarak Antar Kategori pada Setiap Variabel.....	65
5. Perhitungan Manual Mendapatkan <i>Root Node</i> pada <i>CART</i>	66

BAB I

PENDAHULUAN

A. Latar Belakang

Pembangunan nasional bertujuan untuk meningkatkan kesejahteraan masyarakat. Indonesia sebagai negara berkembang melakukan banyak perubahan untuk mendukung pembangunan nasional. Hal tersebut bertujuan agar terciptanya stabilitas nasional, iklim investasi yang baik, dan pertumbuhan ekonomi yang tinggi (Pratiwi dan Zain, 2014). Pertumbuhan ekonomi yang stabil dan berkelanjutan penting untuk pembangunan nasional karena dapat meningkatkan lapangan pekerjaan, pendapatan yang lebih tinggi bagi individu dan rumah tangga, peningkatan standar hidup, dan kemampuan untuk menyediakan layanan publik yang lebih baik seperti pendidikan dan kesehatan. Salah satu penentu kesuksesan suatu perekonomian di bidang sumber daya manusia adalah ketersediaan tenaga kerja (Fajriati, 2022).

Tenaga kerja adalah penduduk dalam usia kerja yang mampu melakukan pekerjaan, antara lain mereka yang sudah bekerja, mereka yang sedang mencari pekerjaan, mereka yang bersekolah, dan mereka yang mengurus rumah tangga (Rionga dan Firdaus, 2007). Badan Pusat Statistik (BPS) membagi penduduk yang merupakan tenaga kerja menjadi dua bagian yaitu angkatan kerja dan bukan angkatan kerja. Angkatan kerja merupakan penduduk usia 15 tahun atau lebih yang melakukan pekerjaan untuk memperoleh pendapatan atau keuntungan termasuk mereka yang sudah mempunyai pekerjaan tetapi seminggu yang lalu tidak bekerja.

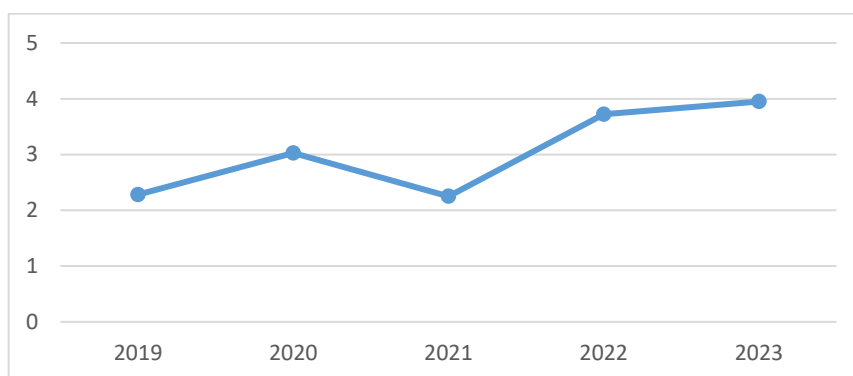
Sedangkan bukan angkatan kerja adalah penduduk usia 15 tahun atau lebih yang tidak bekerja dan tidak mempunyai pekerjaan, karena sedang melakukan kegiatan seperti bersekolah, mengurus rumah tangga, pensiun, dan penderita cacat.

Kemampuan, keterampilan dan keahlian angkatan kerja perlu terus ditingkatkan melalui perencanaan dan program ketenagakerjaan termasuk pelatihan, pemagangan dan pelayanan penempatan tenaga kerja (Charda, 2015). Menurut Pratiwi dan Zain (2014), beberapa masalah yang terkait dengan ketenagakerjaan adalah pengangguran, upah minimum, dan minimnya lapangan pekerjaan. Tingginya pengangguran di suatu daerah dapat diakibatkan melalui rendahnya kemampuan dari angkatan kerja. Untuk mengatasi hal tersebut dapat dipecahkan dengan kebijakan pemerintah yang optimal baik dari sisi pendidikan, perundang-undangan, perluasan lapangan kerja, pelayanan informasi, upah pekerja dan lain sebagainya.

Menurut BPS, pengangguran adalah seseorang yang merupakan angkatan kerja yang telah berusia 15 tahun atau lebih dan tidak sedang menjalani pendidikan formal maupun yang belum memiliki pekerjaan. Pengangguran terdiri dari penduduk usia kerja yang tidak memiliki pekerjaan dan sedang mencari pekerjaan, tidak memiliki pekerjaan dan sedang mempersiapkan usaha, dan tidak memiliki pekerjaan dan tidak mencari pekerjaan, karena merasa tidak mungkin mendapatkan pekerjaan. Selain itu, seseorang yang sudah punya pekerjaan atau usaha tetapi belum memulai pekerjaan atau usaha tersebut juga termasuk pengangguran.

Dalam penelitian Yacoub (2012), Provinsi Kalimantan Barat memiliki tingkat pengangguran yang rendah yaitu 5,44% tetapi memiliki tingkat kemiskinan yang tinggi. Masalah ini penting untuk diteliti karena dapat mempengaruhi pertumbuhan

ekonomi. Masalah ini memiliki keterkaitan dengan daerah Kabupaten Lima Puluh Kota. Kabupaten Lima Puluh Kota memiliki tingkat pengangguran yang rendah dan memiliki tingkat kemiskinan yang tinggi dimana Kabupaten Lima Puluh Kota menduduki posisi 5 kabupaten/kota dengan tingkat kemiskinan yang paling tinggi di Sumatera Barat (BPS, 2023). Berdasarkan data BPS, tingkat Pengangguran di Kabupaten Lima Puluh Kota dapat dilihat pada Gambar 1.



Sumber: BPS

Gambar 1. Tingkat Pengangguran di Kabupaten Lima Puluh Kota pada Tahun 2019-2023

Berdasarkan Gambar 1 dapat dilihat bahwa pengangguran di Kabupaten Lima Puluh Kota mengalami kenaikan pada tahun 2020 dimana kenaikan ini terjadi karena disebabkan oleh pandemi *Covid-19*. Pada tahun 2021, pemerintah banyak melakukan program bantuan yang menciptakan lapangan kerja baru sehingga dapat menurunkan tingkat pengangguran di Kabupaten Lima Puluh Kota (Bappeda Lima Puluh Kota, 2014). Saat pasca pandemi, program tersebut dihentikan dan pengangguran di Lima Puluh Kota mengalami kenaikan yang cukup tinggi yang dirasakan hingga tahun 2023.

Menurut Badan Perencanaan Pembangunan Daerah Kabupaten Lima Puluh Kota (2024), pengangguran di Kabupaten Lima Puluh Kota disebabkan oleh rendahnya kompetensi SDM penduduk untuk memenuhi kebutuhan lapangan

pekerjaan karena rendahnya daya saing SDM di Kabupaten Lima Puluh Kota. Hal ini dipengaruhi oleh kualifikasi pencari kerja belum memenuhi kebutuhan pasar kerja, informasi lowongan kerja masih terbatas, belum optimalnya lembaga pelatihan kerja dimana kapasitas Balai Latihan Kerja (BLK) masih terbatas, dan jenis pelatihan belum sesuai dengan kebutuhan perusahaan. Sementara itu, investasi usaha pada Kabupaten Lima Puluh Kota masih belum merata dan penciptaan lapangan kerja untuk masyarakat lokal masih rendah.

Berdasarkan hasil Survei Angkatan Kerja Nasional (Sakernas) Agustus 2023, Kabupaten Lima Puluh Kota memiliki persentase angkatan kerja berdasarkan pengangguran dan bukan pengangguran dapat dilihat pada Gambar 2.



Sumber: BPS

Gambar 2. Persentase Pengangguran dan Bukan Pengangguran di Kabupaten Lima Puluh Kota Tahun 2023

Berdasarkan Gambar 2 dapat dilihat bahwa, terdapat perbedaan yang cukup besar dari jumlah angkatan kerja yang pengangguran dengan angkatan kerja yang bukan pengangguran, dimana jumlah angkatan kerja yang bukan pengangguran jauh lebih banyak dari pada jumlah angkatan kerja yang pengangguran. Jumlah angkatan kerja yang bukan pengangguran adalah 1029 orang sedangkan jumlah angkatan kerja yang pengangguran adalah 283 orang. Kondisi dimana suatu kelompok kelas memiliki jumlah data yang jauh berbeda dibandingkan dengan

kelas lainnya dikatakan data tidak seimbang (Barro dkk., 2013). Perbedaan jumlah data yang besar antar kelas dapat mengakibatkan model klasifikasi sering tidak dapat memprediksikan kelas minoritas dengan tepat sehingga banyak data tes yang seharusnya berada pada kelas minoritas diprediksikan salah oleh model klasifikasi (Sun dkk., 2009).

Kondisi data tidak seimbang menyulitkan metode klasifikasi dalam melakukan fungsi generalisasi pada proses *machine learning* dimana terjadinya klasifikasi yang menunjukkan performa yang sangat buruk ketika bekerja pada data dengan kelas yang sangat tidak seimbang (Siringoringo, 2018). Jika dataset memiliki rasio ketidakseimbangan 55:45 atau 60:40, maka tidak perlu mengatasi data tidak seimbang, sedangkan jika rasio ketidakseimbangan 80:20 atau 90:10, maka dapat dikatakan data sangat tidak seimbang, sehingga diperlukan teknik penyeimbangan untuk menyeimbangkan data (Shaw dkk., 2021). Jika bekerja pada data tidak seimbang, hampir semua algoritma klasifikasi akan menghasilkan akurasi yang jauh lebih tinggi untuk kelas mayoritas daripada kelas minoritas karena model akan menghasilkan klasifikasi kelas mayor yang lebih banyak dibandingkan kelas minor (Chawla dkk., 2002).

Kondisi ketidakseimbangan data perlu untuk ditangani sehingga memperbaiki performa model klasifikasi. Terdapat berbagai cara untuk menangani data tidak seimbang, salah satunya dengan menggunakan *Synthetic Minority Oversampling Technique (SMOTE)*. Teknik *SMOTE* merupakan salah satu metode penanganan data tidak seimbang dengan membangkitkan data buatan untuk kelas data minoritas sehingga proporsi kelas data mayor dan minor menjadi lebih seimbang. Kelebihan dari metode *SMOTE* secara umum adalah tidak menyebabkan adanya informasi

yang hilang, menghindari terjadinya *overfitting*, membangun wilayah keputusan yang lebih besar, serta mampu meningkatkan akurasi prediksi kelas minoritas. *SMOTE* menyediakan sampel pada kelas minoritas yang lebih terkait untuk dipelajari oleh pengklasifikasi, sehingga pengklasifikasi memiliki cakupan yang lebih besar dalam mempelajari kelas minoritas (Wijayanti dkk., 2021). Hal tersebut menyebabkan pendekatan *SMOTE* mampu meningkatkan nilai akurasi pengklasifikasi pada kelas minoritas, jika dibandingkan *oversampling* dengan duplikasi (Chawla dkk., 2002).

Penelitian yang dilakukan oleh Aritonang dkk. (2016) tentang klasifikasi status kerja pada angkatan kerja Kota Semarang tahun 2014 menggunakan metode *CHAID* dan *CART*, hasil penelitian ini mendapatkan nilai akurasi metode *CHAID* dan *CART* sebesar 72,63% dan 72,79%. Hasil ini menunjukkan metode *CART* lebih baik digunakan untuk klasifikasi status kerja di Kota Semarang. Faktor yang mempengaruhi pengangguran pada angkatan kerja di Kota Semarang yaitu status dalam keluarga, jenis kelamin, usia, status perkawinan, tingkat pendidikan dan pengalaman kerja.

Penelitian yang dilakukan oleh Pratiwi dan Zain (2014) tentang klasifikasi pengangguran terbuka menggunakan *CART* di Provinsi Sulawesi Utara, hasil penelitian ini mendapatkan ketetapan klasifikasi pada *CART* sebesar 78,9%. Faktor yang mempengaruhi pengangguran di provinsi Sulawesi Utara yaitu jenis kelamin, pendidikan terakhir, usia, status dalam rumah tangga, dan status perkawinan. Dengan demikian, perlu adanya pengklasifikasian dari angkatan kerja menjadi pengangguran dan bukan pengangguran di Kabupaten Lima Puluh Kota, yang dapat diklasifikasikan dengan menggunakan metode *decision tree*.

Decision tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari *decision tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (Ratniasih, 2014). Salah satu metode *decision tree* untuk mengklasifikasikan angkatan kerja menjadi pengangguran dan bukan pengangguran adalah metode *Classification and Regression Tree (CART)*. Metode *CART* dapat diterapkan pada data berdimensi tinggi dan dapat menyeleksi variabel dan interaksi antar variabel yang paling penting dalam menentukan variabel dependennya.

Menurut Sartono dan Syafitri (2010), metode *CART* memiliki kemampuan dalam memberikan dugaan dengan tingkat kesalahan yang kecil dan kemudahan interpretasi terhadap hasil analisis. *CART* bertujuan untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Selain itu, *CART* digunakan untuk menggambarkan hubungan antara variabel dependen dengan satu atau lebih variabel independen. Hasil klasifikasi ini dapat digunakan untuk mengetahui karakteristik setiap kategori angkatan kerja.

Berdasarkan permasalahan tersebut, maka dilakukan penelitian untuk mengklasifikasikan pengangguran di Kabupaten Lima Puluh Kota menggunakan metode *CART*. Sehingga dilakukan penelitian yang berjudul **“Penanganan Data Tidak Seimbang dengan Algoritma *SMOTE* untuk Klasifikasi Pengangguran di Kabupaten Lima Puluh Kota Menggunakan Metode *CART*”**.

B. Batasan Masalah

Berdasarkan latar belakang masalah yang telah dikemukakan, batasan masalah dalam penelitian ini adalah

1. Banyak faktor yang dapat mempengaruhi pengangguran namun yang menjadi fokus dalam penelitian ini adalah faktor usia, jenis kelamin, status perkawinan, status dalam keluarga, tingkat pendidikan, pengalaman kerja, dan daerah tempat tinggal.
2. Penanganan data tidak seimbang menggunakan algoritma *SMOTE* dan analisis yang digunakan untuk mengklasifikasikan pengangguran di Kabupaten Lima Puluh Kota adalah metode *CART*.

C. Rumusan Masalah

Rumusan masalah yang dibahas pada penelitian ini adalah sebagai berikut.

1. Bagaimana hasil pohon klasifikasi pengangguran dan bukan pengangguran di Kabupaten Lima Puluh Kota tahun 2023 menggunakan metode *CART* sebelum dan sesudah diterapkan *SMOTE*?
2. Apa saja karakteristik dari pengangguran dan bukan pengangguran di Kabupaten Lima Puluh Kota dengan menggunakan *CART* sebelum dan sesudah *SMOTE*?
3. Bagaimana tingkat akurasi, sensitivitas, dan spesifisitas pohon klasifikasi pengangguran dan bukan pengangguran di Kabupaten Lima Puluh Kota tahun 2023 menggunakan *CART* sebelum dan sesudah diterapkan *SMOTE*?

D. Tujuan Penelitian

1. Untuk mengetahui hasil pohon klasifikasi pengangguran dan bukan pengangguran di Kabupaten Lima Puluh Kota tahun 2023 menggunakan metode *CART* sebelum dan sesudah diterapkan *SMOTE*.

2. Untuk mengetahui karakteristik dari pengangguran dan bukan pengangguran di Kabupaten Lima Puluh Kota dengan menggunakan *CART* sebelum dan sesudah *SMOTE*.
3. Untuk mengetahui tingkat akurasi, sensitivitas, dan spesifisitas pohon klasifikasi pengangguran dan bukan pengangguran di Kabupaten Lima Puluh Kota tahun 2023 menggunakan *CART* sebelum dan sesudah diterapkan *SMOTE*.

E. Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian ini adalah sebagai berikut.

1. Bagi peneliti, dapat memberikan pengetahuan baru mengenai penerapan *SMOTE* dalam mengatasi data tidak seimbang untuk klasifikasi pengangguran di Kabupaten Lima Puluh Kota tahun 2023 dengan menggunakan metode *CART*.
2. Bagi pembaca, dapat menambah wawasan dan ilmu pengetahuan khususnya dalam metode yang digunakan dan hasil yang didapatkan untuk mengetahui klasifikasi pengangguran di Kabupaten Lima Puluh Kota tahun 2023.
3. Bagi pemerintah, dapat memberikan sumbangan pemikiran atau sebagai rujukan dalam pembuatan kebijakan dan keputusan yang berhubungan dengan tenaga kerja

BAB V

KESIMPULAN

A. Kesimpulan

Berdasarkan analisis yang telah dilakukan, maka dapat diperoleh Kesimpulan sebagai berikut.

1. Pohon klasifikasi sebelum diterapkan *SMOTE* menghasilkan 4 *terminal node* dengan 3 variabel independen yaitu usia (X_1), pengalaman kerja (X_6), dan tingkat pendidikan (X_5). Pohon klasifikasi menghasilkan 3 *terminal node* yang mengklasifikasikan bukan pengangguran dan 1 mengklasifikasikan pengangguran. Pohon klasifikasi *CART* sesudah diterapkan *SMOTE* menghasilkan 6 *terminal node*, yang terdiri dari 5 variabel independen, yaitu usia (X_1), jenis kelamin (X_2), tingkat pendidikan (X_5) pengalaman kerja (X_6), dan daerah tempat tinggi (X_7). Pohon klasifikasi menghasilkan 3 *terminal node* yang mengklasifikasikan bukan pengangguran dan 3 *terminal node* yang mengklasifikasikan pengangguran.
2. Karakteristik bukan pengangguran pada pohon klasifikasi *CART* sebelum diterapkan *SMOTE* yaitu:
 - a. Berkategori usia produktif atau usia tidak produktif.
 - b. Berkategori usia muda dan memiliki pengalaman kerja
 - c. Berkategori usia muda, tidak memiliki pengalaman kerja, dan tingkat pendidikan tamat SLTA ke atas.

Angkatan kerja akan menjadi pengangguran jika berkategori usia muda, tidak memiliki pengalaman kerja, dan tingkat pendidikan tamat SLTP ke bawah.

Karakteristik bukan pengangguran pada pohon klasifikasi *CART* sesudah diterapkan *SMOTE* yaitu:

- a. Jika berkategori usia produktif atau usia tidak produktif dan berjenis kelamin laki-laki.
- b. Jika berkategori usia produktif atau usia tidak produktif, berjenis kelamin perempuan, dan tingkat pendidikannya tamat akademi atau perguruan tinggi.
- c. Jika berkategori usia produktif atau usia tidak produktif, berjenis kelamin perempuan, tingkat pendidikannya tamat SLTA ke bawah, memiliki pengalaman kerja, dan tinggal di daerah pedesaan.

Karakteristik pengangguran pada pohon klasifikasi terbaik yaitu:

- a. Berkategori usia muda.
 - b. Berkategori usia produktif atau usia tidak produktif, berjenis kelamin perempuan, tingkat pendidikannya tamat SLTA ke bawah, memiliki pengalaman kerja, dan tinggal di daerah perkotaan.
 - c. Berkategori usia produktif atau usia tidak produktif, berjenis kelamin perempuan, tingkat pendidikannya tamat SLTA ke bawah, tidak memiliki pengalaman kerja.
3. Penerapan metode *SMOTE* pada klasifikasi *CART* dapat meningkatkan nilai akurasi pada kelas minoritas sebesar 33% sehingga menaikkan nilai *AUC* sebesar 7%. Hal ini menunjukkan bahwa dengan melakukan

penanganan data tidak seimbang mampu memperbaiki kinerja klasifikasi. Maka dapat dikatakan bahwa metode *CART* sesudah diterapkan *SMOTE* memberikan kinerja yang lebih baik sehingga metode *CART* sesudah diterapkan *SMOTE* merupakan metode terbaik dalam mengklasifikasikan pengangguran di Kabupaten Lima Puluh Kota.

B. Saran

Penelitian ini menghasilkan ketetapan hasil klasifikasi pengangguran yang cukup baik, sehingga penelitian selanjutnya dapat mencobakan penerapan data tidak seimbang metode *ensemble oversampling* dan *undersampling*. Dimana metode tersebut melakukan pembuatan data baru sekaligus menghilangkan beberapa data agar proporsinya seimbang. Apabila setelah dicobakan menggunakan metode *ensemble* namun masih belum meningkatkan nilai ketetapan hasil klasifikasi pada penelitian ini, maka dapat diduga bahwa metode *CART* masih dikategorikan cukup baik dalam mengklasifikasikan pengangguran.

DAFTAR PUSTAKA

- Adioetomo, S. M. dan Samosir, O. B. (2010). *Dasar-dasar Demografi*, Jakarta: Salemba Empat.
- Aritonang, N. E., Rusgiyono, A., dan Rahmawati, R. (2016). Klasifikasi Status Kerja Pada Angkatan Kerja Kota Semarang Tahun 2014 Menggunakan Metode CHAID dan CART. *Jurnal Gaussian*, Vol. 5, No. 1.
- Badan Perencanaan Pembangunan Daerah Kabupaten Lima Puluh Kota (2024). *Rancangan Awal Rencana Pembangunan Jangka Panjang (RPJPD) Kabupaten Lima Puluh Kota Tahun 2025-2045*. Lima Puluh Kota, Indonesia: Badan Perencanaan Pembangunan Daerah.
- Badan Pusat Statistik (BPS) (2023). *Keadaan Angkatan Kerja di Indonesia*, Indonesia: Badan Pusat Statistik.
- Barro, R. A., Sulvianti, I. D., dan Afendi, F. M. (2013). Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu. *Xplore: Journal of Statistics*, Vol. 1, No. 1.
- Breiman, L., Friedman, J. H., Olshen, R. A., dan Stone, C. J. (1984). *Classification And Regression Trees*, New York: Chapman and Hall.
- Charda, U. S. (2015). Karakteristik Undang-undang Ketenagakerjaan Dalam Perlindungan Hukum Terhadap Tenaga Kerja. *Jurnal Wawasan Hukum*, Vol. 32, No. 1.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., dan Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique, *Journal Of Artificial Intelligence research*, Vol. 16, No. 2.
- Cost, S. dan Salzberg, S. (1993). A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features, *Machine Learning*, 10.
- Desanti, G. dan Ariusni (2022). Pengaruh Umur, Jenis Kelamin, Jam Kerja, Status Pekerjaan dan Pendidikan Terhadap Pendapatan Tenaga Kerja di Kota Padang, *Jurnal Kajian Ekonomi dan Pembangunan*, Vol. 3, No. 4.
- Ehrenberg, R.G dan Smith, R. S. (1994). *Modern Labor Economics*. Fifth Edition. New York, NY: Harper Collins.
- Fajriati, Y. R. (2022). Pengklasifikasian Status Kerja pada Angkatan Kerja di Kabupaten Tanah Datar Menggunakan Metode CART dan Metode CHAID. *Journal Of Mathematics UNP*, Vol. 7, No. 3.
- Fawcett, T. (2005). An Introduction to ROC analysis. *Pattern Recognition Letters*, 27.

- Harfina, D. S. (2007). Faktor-faktor yang Mempengaruhi Pengangguran Terselubung di Pedesaan Jawa Tengah Analisis Data Sakernas 2007. *Jurnal Kependudukan Indonesia*, Vol. 4, No. 1.
- Hasanah, U., Zulham, T., Mahrizal, dan Affandi (2021). Pengaruh Migrasi Masuk dan Upah Minimum Provinsi Terhadap Pengangguran di Kota Banda Aceh. *Jurnal Ekombis*, Vol. 7, No. 1.
- Herawati, N. dan Sasana, H. (2013). Analisis Pengaruh Pendidikan, Upah Pengalaman Kerja, Jenis Kelamin dan Umur Terhadap Produktivitas Tenaga Kerja Industri Shutllecock Kota Tegal. *Diponegoro Journal of Economics*, Vol. 2, No. 4.
- Komalasari, W. B. (2007). Metode Pohon Regresi Untuk Eksploratori Data Dengan Peubah yang Banyak dan Kompleks. *Informatika Pertanian*, Vol. 16, No. 1.
- Lewin, R. J., (2000), "An Introduction to Classification and Regression Tree (CART) Analysis", *Presented at Annual Meeting of the Society for Academic Emergency Medicine*, San Francisco, California, 2000.
- Ngadi. (2005). Pengangguran Terbuka dan Setengah Pengangguran di Indonesia Periode 1996-2004: Konsep, Isu dan Implikasi Kebijakan. *Warta Demografi*. Vol. 35, No. 4.
- Prasodjo, I. (2018). Dampak Urbanisasi Bagi Pembangunan Manusia 2010-2016 Studi Kasus: Jakarta, Surabaya dan Medan. *Jurnal Ekonomi*, Vol. 23, No. 3.
- 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.
- Ramadhanti, D. V., Santoso, R., dan Widiharih, T. (2022). Perbandingan SMOTE dan ADASYN pada Data Imbalance Untuk Klasifikasi Rumah Tangga Miskin Kabupaten Temanggung dengan Algoritma K-Nearest Neighbor. *Jurnal Gaussian*, Vol. 11, No. 4.
- Ratniasih, N. L. (2014). Konversi Data Training Tentang Pemilihan Kelas Menjadi Bentuk Pohon Keputusan Dengan Teknik Klasifikasi. *Eksplora Informatika* Vol. 3, No. 2.
- Rionga, MT., dan Firdaus, Y. (2007). *Tenaga Kerja*, Bandung: Alfabeta.
- Saputro, I. W. dan Sari, B. W. (2019). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Citec Journal*, Vol. 6, No. 1.
- Sari, R. V., Firdausi, F., dan Azhar, Y. (2020). Perbandingan Prediksi Kualitas Kopi Arabisa dengan Menggunakan Algoritma SGD, Random Forest, dan Naïve Bayes. *Jurnal Pendidikan Informatika*, Vol. 4, No. 2.
- Sartono, B., dan Syafitri, U.D. (2010). Metode Pohon Gabungan: Solusi Pilihan Untuk Mengatasi Kelemahan Pohon Regresi dan Klasifikasi Tunggal. *Forum Statistika dan Komputasi*, Vol. 4, No. 1.

- Shaw, S. S., Ahmed, S., Malakar, S., Hernandez, L. G., Abraham, A., dan Sarkar, R. (2021). Hybridization of Ring Theory-based Evolutionary Algorithm and Particle Swarm Optimization to Solve Class Imbalance Problem. *Springer* Vol. 7.
- Simanjuntak, P. J. (2001). *Pengantar Ekonomi Sumber Daya Manusia*. Jakarta: Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia.
- Siringoringo, R. (2018). Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor. *Jurnal ISD* Vol. 3. No.1.
- Soeprobo, T. B. (2002). *Indonesia Youth Employment*, Prepared for ILO/Japan Tripartite Regional Meeting on Youth Employment in Asia and the Pacific Bangkok, 27 February-1 March 2002.
- Soetomo, H. (1984). *Pengangguran di Kota, Suatu Analisis terhadap Pemuda dan Golongan Kerja dalam Angkatan Kerja Indonesia*. Jakarta: Rajawali.
- Sukirno, S. (1994). *Pengantar Teori Ekonomi*. Jakarta: Raja Grafindo Persada.
- Sumartini, S. H. dan Purmani, S. H. (2015). Penggunaan Metode Classification and Regression Tree (CART) Untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya. *Jurnal Sains dan Seni ITS*, Vol. 4, No. 2.
- Sun, Y., Wong, A. K. C., dan Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*. Vol. 23, No. 4.
- Wijayanti, N. P. Y. T., Kencana, E. N., dan Sumarjaya, I. W. (2021). SMOTE: Potensi dan Kekurangannya Pada Survei. *E-Jurnal Matematika* Vol. 10, No. 4.