

**PENERAPAN SMOTE PADA METODE CART UNTUK MENGATASI  
*IMBALANCED DATA* DALAM PENGKLASIFIKASIAN STATUS  
KERJA PADA ANGKATAN KERJA DI KOTA PADANG**

**SKRIPSI**



**Oleh  
ANDINI YULIANTI  
NIM 18337010**

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2022**

**PENERAPAN SMOTE PADA METODE CART UNTUK MENGATASI  
*IMBALANCED DATA* DALAM PENGKLASIFIKASIAN STATUS  
KERJA PADA ANGKATAN KERJA DI KOTA PADANG**

**SKRIPSI**

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



Oleh  
**ANDINI YULIANTI**  
**NIM 18337010**

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2022**

## PERSETUJUAN SKRIPSI


### PENERAPAN SMOTE PADA METODE CART UNTUK MENGATASI *IMBALANCED DATA* DALAM PENGKLASIFIKASIAN STATUS KERJA PADA ANGKATAN KERJA DI KOTA PADANG

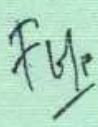
Nama : Andini Yulianti  
NIM : 18337010  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 26 Agustus 2022

Mengetahui:  
Ketua Departemen Statistika

Disetujui Oleh:  
Pembimbing

  
Dr. Dony Permana, M.Si  
NIP. 197501272006041001

  
Fadhilah Fitri, S.Si., M.Stat  
NIP. 199403142022032015

## PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Andini Yulianti  
NIM : 18337010  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

### PENERAPAN SMOTE PADA METODE CART UNTUK MENGATASI *IMBALANCED DATA* DALAM PENGKLASIFIKASIAN STATUS KERJA PADA ANGKATAN KERJA DI KOTA PADANG



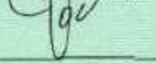
Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi  
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Negeri Padang

Padang, 26 Agustus 2022

#### Tim Penguji

	Nama
Ketua	: Fadhilah Fitri, S.Si., M. Stat
Anggota	: Dra. Nonong Amalita, M.Si
Anggota	: Dodi Vionanda, M. Si., Ph.D

#### Tanda Tangan

  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_

## SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan di bawah ini:

Nama : Andini Yulianti  
NIM/Tahun Masuk : 18337010/2018  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, Skripsi saya dengan judul **“Penerapan SMOTE pada Metode CART untuk Mengatasi *Imbalanced Data* dalam Pengklasifikasian Status Kerja pada Angkatan Kerja di Kota Padang”** 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,  
Ketua Departemen Statistika,

  
Dr. Dony Permana, M.Si  
NIP. 97501272006041001

Padang, 26 Agustus 2022

Saya yang menyatakan,



Andini Yulianti  
NIM. 18337010

# **Penerapan SMOTE pada Metode CART untuk Mengatasi *Imbalanced Data* dalam Pengklasifikasian Status Kerja pada Angkatan Kerja di Kota Padang**

**Andini Yulianti**

## **ABSTRAK**

Pengangguran merupakan masalah kompleks yang dihadapi oleh negara berkembang seperti Indonesia. Tingginya angka pengangguran di Indonesia berdampak pada kemiskinan sehingga pemerintah berupaya untuk melakukan pembangunan ekonomi. Kota Padang merupakan kota yang memiliki tingkat pengangguran tertinggi di Sumatera Barat dari tahun 2013 sampai 2021. Pengembangan kota smart dan pengidentifikasian faktor yang mempengaruhi pengangguran merupakan salah satu upaya untuk mengurangi pengangguran.

Penelitian ini menggunakan metode CART untuk mengetahui faktor-faktor yang mempengaruhi jumlah angkatan kerja di kota Padang. Keuntungan metode CART adalah mudah untuk diinterpretasikan hasil analisisnya, namun keakuratan pohon klasifikasi rendah karena ketidakseimbangan data. Oleh karena itu, penelitian ini menggunakan metode SMOTE untuk mengatasi permasalahan tersebut.

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan SMOTE dapat meningkatkan akurasi metode CART yang diterapkan pada data tidak seimbang yaitu sebesar sebesar 43,1%. Pohon klasifikasi optimal terbentuk dari 8 simpul terminal dan melibatkan 4 peubah penjelas yang terdiri dari status perkawinan, tingkat pendidikan, jenis kelamin dan usia, 5 simpul terminal yang mengklasifikasikan angkatan kerja berkategori bekerja dan 3 simpul terminal yang mengklasifikasikan angkatan kerja berkategori pengangguran.

**Kata Kunci:** Angkatan Kerja, Padang, CART, SMOTE

# **The SMOTE Application of CART Methods for Coping Imbalanced Data in Classifying Status Work on Labor Force in the City of Padang**

**Andini Yulianti**

## **ABSTRACT**

Unemployment is a complex problem faced by developing countries, one of which is Indonesia. The high unemployment rate in Indonesia has an impact on poverty so that the government seeks to carry out economic development. The city of Padang is a city that has the highest unemployment rate in West Sumatra from 2013 to 2021. The development of a smart city and identification of factors that influence unemployment is one of the efforts to reduce unemployment.

This study uses the CART method to determine the factors that influence the number of the workforce in the city of Padang. The advantage of the CART method is that it is easy to interpret the results of the analysis, but the accuracy of the classification tree is low due to data imbalance. Therefore, this study uses the SMOTE method to overcome these problems.

Based on the analysis that has been done, it can be concluded that SMOTE can increase the accuracy of the CART method applied to unbalanced data by 43.1%. The optimal classification tree is formed from 8 terminal nodes and involves 4 explanatory variables consisting of marital status, education level, gender and age, 5 terminal nodes which classify the labor force into the working category and 3 terminal nodes which classify the labor force into the unemployed category.

**Keywords:** Workforce, Padang, CART, SMOTE

## KATA PENGANTAR



Bismillahirrahmanirrahim, *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 **“Penerapan SMOTE pada Metode CART untuk Mengatasi *Imbalanced Data* dalam pengklasifikasian Status Kerja pada Angkatan Kerja di Kota Padang”** 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 Fadhilah Fitri, S.Si., M.Stat, Pembimbing Skripsi yang telah memberikan bimbingan dan dorongan dalam menyelesaikan Skripsi ini, sekaligus sebagai tenaga pengajar Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.



2. Bapak Zamahsary Martha, S.Si., M.Si, Selaku Pembimbing Akademik yang telah memberikan bimbingan dan dorongan selama masa perkuliahan, sekaligus sebagai tenaga pengajar dan Kepala Labor Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
3. Bapak Dodi Vionanda, M.Si., Ph.D, Dosen Penguji yang banyak memberikan masukan dan arahan demi kesempurnaan Skripsi ini, sekaligus Koordinator Prodi Diploma Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
4. Ibu Dra. Nonong Amalita, M.Si, Dosen Penguji yang banyak memberikan masukan dan arahan demi kesempurnaan Skripsi ini, sekaligus Sekretaris Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
5. Bapak dan Ibu Dosen Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
6. Kedua Orang Tua, abang dan adik yang selalu mendoakan, memberi semangat, dan nasihat dalam menyelesaikan Skripsi ini.
7. Serta semua sahabat dan rekan-rekan Statistika 2018 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, Agustus 2022

Andini Yulianti

## DAFTAR ISI

<b>ABSTRAK .....</b>	<b>i</b>
<b>KATA PENGANTAR.....</b>	<b>iii</b>
<b>DAFTAR ISI.....</b>	<b>vi</b>
<b>DAFTAR TABEL .....</b>	<b>viii</b>
<b>DAFTAR GAMBAR.....</b>	<b>ix</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>x</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
A. Latar Belakang Masalah.....	1
B. Batasan Masalah.....	5
C. Rumusan Masalah .....	5
D. Tujuan Penelitian .....	6
E. Manfaat Penelitian .....	6
<b>BAB II KERANGKA TEORITIS .....</b>	<b>7</b>
A. Kajian Teori .....	7
B. Kajian Terdahulu.....	21
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>23</b>
A. Jenis Penelitian.....	23
B. Jenis dan Sumber Data .....	23
C. Populasi dan Sampel .....	23
D. Variabel-variabel dalam Penelitian.....	23
E. Teknik Analisis Data.....	24

<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>26</b>
A. Hasil .....	26
B. Pembahasan.....	40
<b>BAB V KESIMPULAN .....</b>	<b>42</b>
A. Kesimpulan .....	42
B. Saran.....	43
<b>DAFTAR PUSTAKA .....</b>	<b>45</b>
<b>LAMPIRAN.....</b>	<b>48</b>

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. Struktur Data Ketepatan Klasifikasi .....	18
2. Variabel Dalam Penelitian .....	24
3. Distribusi Data dengan Menggunakan SMOTE .....	31
4. Data Calon Simpul Kiri dan Calon Simpul Kanan .....	32
5. Hasil Perhitungan Probabilitas Simpul .....	33
6. Hasil Perhitungan Indeks Gini .....	34
7. Hasil Perhitungan Goodness of Split .....	35
8. Ketepatan Klasifikasi CART tanpa SMOTE .....	36
9. Struktur Data Ketepatan Klasifikasi Metode CART dengan SMOTE .....	39
10. Perbandingan CART tanpa SMOTE dan dengan SMOTE .....	39

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
1. Jumlah Angkatan Kerja Bekerja dan Pengangguran di Kota Padang Tahun 2021.....	4
2. Bagan Penduduk dan Angkatan Kerja .....	7
3. Pohon Klasifikasi CART .....	15
4. Deskripsi Data Status Kerja Tahun 2021 .....	27
5. Deskripsi Usia .....	27
6. Deskripsi Jenis Kelamin .....	28
7. Deskripsi Status Perkawinan.....	28
8. Deskripsi Tingkat Pendidikan.....	29
9. Deskripsi Kartu Prakerja.....	29
10. Persentase Banyaknya Variabel Dependen Setelah SMOTE .....	31
11. Pohon Klasifikasi tanpa SMOTE.....	36
12. Pohon Klasifikasi dengan SMOTE.....	37

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
1. Data Penelitian .....	48
2. Hasil Sebelum SMOTE.....	49
3. Perhitungan Pembentukan Pohon Klasifikasi pada Metode CART Setelah SMOTE .....	50
4. Syntax Metode CART Setelah SMOTE .....	50
5. Karakteristik Bekerja dan Pengangguran pada Pohon Klasifikasi CART dengan SMOTE .....	52
6. Karakteristik Bekerja dan Pengangguran pada Pohon Klasifikasi CART tanpa SMOTE .....	53
7. Data Setelah dilakukan SMOTE.....	53

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang Masalah**

Masalah ketenagakerjaan menjadi salah satu perhatian utama di setiap negara, khususnya di negara yang sedang berkembang. Indonesia sebagai negara berkembang tidak terlepas dari masalah tersebut. Masalah ketenagakerjaan yang dihadapi oleh Indonesia adalah kurangnya ketersediaan lapangan pekerjaan, kelebihan tenaga kerja, dan persebaran tenaga kerja yang tidak merata. Hal ini disebabkan karena pertumbuhan angkatan kerja lebih tinggi dibandingkan dengan pertumbuhan kesempatan kerja yang ada. Jika permasalahan ini tidak segera diatasi akan menyebabkan meningkatnya jumlah pengangguran.

Tenaga kerja merupakan semua penduduk yang telah mencapai batas usia kerja tertentu. Menurut Badan Pusat Statistik (BPS), tenaga kerja merupakan setiap penduduk yang masuk ke dalam usia kerja (15 tahun keatas). Dimana Indonesia memiliki batasan usia minimal kerja 15 tahun, tetapi tidak memiliki batasan usia maksimal. Hal ini sudah diatur dalam Undang-Undang Nomor 25 Tahun 1997 tentang ketenagakerjaan. Akan tetapi usia paling produktif seseorang dalam bekerja adalah ketika berumur 25 tahun keatas. Oleh karena itu mereka yang sudah mencapai usia pensiun dan masih tetap bekerja disebut tenaga kerja (Simanjuntak, 2001)

Penduduk Indonesia yang memiliki usia kerja pada bulan Agustus 2021 sebanyak 206,71 juta jiwa dan 140,15 juta jiwa adalah angkatan kerja, naik



sebanyak 1,93 juta jiwa dibanding Agustus 2020 (BPS, 2021). Jumlah angkatan kerja yang meningkat setiap tahun tidak seiring dengan lapangan pekerjaan yang tersedia, sehingga mengakibatkan tingginya tingkat pengangguran.

BPS membagi penduduk yang merupakan tenaga kerja menjadi dua bagian yaitu angkatan kerja dan bukan angkatan kerja. Angkatan kerja terdiri dari bekerja dan tidak bekerja (pengangguran) (Dinas Tenaga Kerja, 2019). 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 pengangguran karena cuti, sakit, mogok, mangkir, menunggu panen dan lain-lain. Sedangkan bukan angkatan kerja adalah penduduk usia 15 tahun atau lebih yang pengangguran dan tidak mempunyai pekerjaan, karena sedang melakukan kegiatan seperti bersekolah, mengurus rumah tangga, pensiun, dan penderita cacat (BPS, 2021).

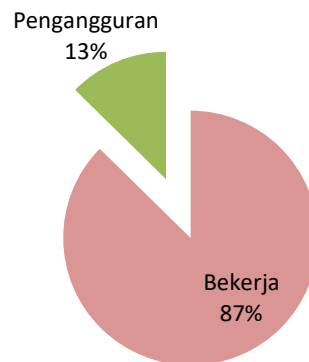
Klasifikasi merupakan dasar untuk proses prediksi, yaitu dengan cara mempelajari karakteristik data untuk membentuk suatu model (Rahmayanti, *et al.* 2021). Menurut Prasetyo (2012) klasifikasi adalah penilaian objek data dengan cara memasukkan data kedalam kelas tertentu dari sejumlah kelas yang tersedia. Pohon keputusan (*Decision Tree*) merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) dimana setiap *node* mempresentasikan atribut, cabangnya mempresentasikan nilai dari atribut, dan daun mempresentasikan kelas. *Node* yang paling atas dari *Decision Tree* disebut *root*.

CART (*Classification And Regression Tree*) merupakan salah satu dari 10 algoritma penambangan data terbaik (Wu, X dkk. 2007). Metode CART dapat menyeleksi peubah dan interaksi antar peubah yang paling penting dalam menentukan peubah responnya. Metode CART sudah diakui tingkat keberhasilannya untuk menganalisis algoritma berdasarkan tingkat akurasi dan waktu tempuh yang diperlukan.

Pada penelitian ini dilakukan analisis yang bertujuan untuk membandingkan hasil klasifikasi angkatan kerja yang bekerja dan pengangguran sebelum dan sesudah diterapkan SMOTE pada metode CART serta untuk mengetahui faktor-faktor yang berpengaruh terhadap angkatan kerja yang bekerja dan pengangguran di Kota Padang tahun 2021. Penelitian sebelumnya yang dilakukan oleh Pratiwi dan Zain (2014) menunjukkan bahwa faktor usia, jenis kelamin, tingkat pendidikan, status perkawinan, dan status dalam rumah tangga berpengaruh secara signifikan terhadap peluang terjadinya pengangguran, namun tidak demikian untuk pengalaman pelatihan kerja dan klasifikasi tempat tinggal. Pada penelitian ini akan digunakan 2 variabel dependen yaitu angkatan kerja yang bekerja dan pengangguran, dengan 5 variabel independen yaitu usia, jenis kelamin, status perkawinan, tingkat pendidikan, dan kartu prakerja.

CART memiliki kemampuan dalam memberikan kemudahan untuk menginterpretasikan hasil analisis dan memberikan dugaan dengan tingkat kesalahan yang kecil (Sartono dan Syafitri, 2010). Tujuan utama CART adalah untuk mendapatkan sekelompok data yang homogen sebagai penciri dari suatu pengklasifikasian. Hasil klasifikasi diharapkan dapat digunakan untuk mengetahui

karakteristik setiap kategori angkatan kerja. Perbandingan jumlah responden angkatan kerja yang berstatus bekerja dan pengangguran di Kota Padang tahun 2021 dalam persen dapat dilihat pada Gambar 1.



Sumber: BPS Kota Padang

**Gambar 1. Jumlah Angkatan Kerja Bekerja dan Pengangguran di Kota Padang Tahun 2021**

Berdasarkan Gambar 1, data jumlah angkatan kerja yang bekerja dan pengangguran tidak seimbang (*imbalance*), dimana jumlah data angkatan kerja yang berstatus bekerja jauh lebih banyak dibanding jumlah angkatan kerja yang berstatus pengangguran. 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 (Y. Sun, *et al*, 2011).

Ketidakseimbangan jumlah amatan pada masing-masing kelas peubah dependen, yakni terdapat kelas yang lebih mendominasi (*mayor*) dan kelas yang jumlah amatannya jauh lebih sedikit (*minor*) menyebabkan kesalahan dalam klasifikasi pada kelas *minor*, atau pada penelitian ini adalah pengangguran. Oleh karena itu, salah satu alternatif untuk meningkatkan kinerja klasifikasi dan

prediksi tersebut adalah dengan menerapkan metode resampling yaitu *Synthetic Minority Oversampling Technique* (SMOTE) untuk menangani data tidak seimbang (*imbalance class*). SMOTE dapat meningkatkan akurasi pengklasifikasian untuk kelas *minor* (Chawla *et al.* 2002). Dengan menggunakan model klasifikasi *Decision Tree* dan algoritma SMOTE sebagai penyeimbang data dari kelas minoritas, diharapkan dapat membantu dan menjadi solusi untuk meminimalisir terjadinya pengangguran di Kota Padang. Berdasarkan permasalahan di atas, maka penelitian ini diberi judul **“Penerapan SMOTE Pada Metode CART Untuk Mengatasi *Imbalanced Data* Dalam Pengklasifikasian Status Kerja pada Angkatan Kerja Di Kota Padang”**.

## **B. Batasan Masalah**

Pada penelitian ini, perlu adanya batasan masalah dengan maksud untuk memperjelas arah dari suatu masalah agar tidak menimbulkan kekeliruan. Batasan masalah pada penelitian ini mencakup faktor usia, jenis kelamin, status perkawinan, tingkat pendidikan dan kartu prakerja di Kota Padang tahun 2021.

## **C. Rumusan Masalah**

Adapun rumusan masalah dalam penelitian ini adalah.

1. Bagaimana hasil pohon klasifikasi status kerja pada angkatan kerja di Kota Padang tahun 2021 menggunakan metode CART sebelum dan sesudah diterapkan SMOTE?

2. Bagaimana tingkat akurasi pohon klasifikasi status kerja pada angkatan kerja di Kota Padang tahun 2021 menggunakan metode CART sebelum dan sesudah diterapkan SMOTE?

#### **D. Tujuan Penelitian**

Tujuan dalam penelitian ini adalah sebagai berikut.

1. Untuk mengetahui hasil pohon klasifikasi status kerja pada angkatan kerja di Kota Padang tahun 2021 menggunakan metode CART sebelum dan sesudah diterapkan SMOTE.
2. Untuk mengetahui tingkat akurasi pohon klasifikasi status kerja pada angkatan kerja di Kota Padang tahun 2021 menggunakan metode CART sebelum dan sesudah diterapkan SMOTE.

#### **E. Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini adalah.

1. Penulis, dapat menambah ilmu pengetahuan dan pemahaman terhadap analisis CART dan SMOTE.
2. Bagi pemerintahan, dapat memberikan sumbangan pemikiran atau sebagai rujukan dalam pembuatan kebijakan dan keputusan yang berhubungan dengan tenaga kerja.
3. Peneliti selanjutnya, sebagai referensi untuk mengembangkan dan memperluas cakupan hasil penelitian.