



TUGAS AKHIR - SS141501

**PENGLASIFIKASIAN PARTISIPASI KEGIATAN
EKONOMI LANSIA DI JAWA TIMUR DENGAN
MENGUNAKAN METODE CART (*CLASSIFICATION
AND REGRESSION TREES*)**

PITRI ARISKA SUSILOWATI
NRP 1313 105 015

Dosen Pembimbing
Dr. Wahyu Wibowo, S.Si, M.Si

Program Studi S1 Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2015



FINAL PROJECT- SS141501

**CLASSIFICATION OF PARTICIPATION ECONOMIC
ACTIVITIES ELDERLY IN EAST JAVA USING CART
(CLASSIFICATION AND REGRESSION TREES)**

**PITRI ARISKA SUSILOWATI
NRP 1313 105 015**

**Supervisor
Dr. Wahyu Wibowo, S.Si, M.Si**

**Undergraduate Program Departement Of Statistics
Faculty of Mathematics and Natural Science
Sepuluh Nopember Institute Of Technology
Surabaya 2015**

LEMBAR PENGESAHAN

PENGLASIFIKASIAN PARTISIPASI KEGIATAN EKONOMI LANSIA DI JAWA TIMUR DENGAN MENGUNAKAN METODE CART (*CLASSIFICATION AND REGRESSION TRESS*)

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada

Program Studi Sarjana (S1) Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

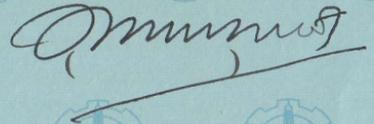
Oleh :

PITRI ARISKA SUSILOWATI

NRP. 1313 105 015

Disetujui oleh Pembimbing Tugas Akhir :

Dr. Wahyu Wibowo, S.Si, M.Si
NIP. 19740328 199802 1 001



Mengetahui
Ketua Jurusan Statistika FMIPA-ITS

Dr. Muhammad Mashuri, MT.
NIP. 19620408 198701 1 001

SURABAYA, Juli 2015

Pengklasifikasian Partisipasi Kegiatan Ekonomi Lansia di Jawa Timur dengan Menggunakan Metode CART (*Classification And Regression Trees*)

Nama Mahasiswa : Pitri Ariska Susilowati
NRP : 1313 105 015
Jurusan : Statistika FMIPA-ITS
Pembimbing : Dr. Wahyu Wibowo, S.Si, M.Si

Abstrak

*Kegiatan ekonomi lansia yaitu lansia yang melakukan kegiatan dengan tujuan mendapat keuntungan, minimal satu jam selama seminggu. Lansia yaitu penduduk yang berumur 60 tahun keatas. Partisipasi kegiatan ekonomi lansia dibedakan menjadi status lansia bekerja dan tidak bekerja. Provinsi Jawa Timur merupakan salah satu provinsi yang memiliki persentase lansia bekerja yang tinggi. Permasalahannya adalah bagaimana klasifikasi lansia di provinsi Jawa Timur yang bekerja dan tidak bekerja berdasarkan faktor-faktor yang mempengaruhinya dan mengetahui faktor-faktor yang berpengaruh terhadap partisipasi kegiatan ekonomi lansia. CART (*Classification And Regression Trees*) merupakan salah satu metode yang digunakan untuk pembentukan pohon klasifikasi. Data yang digunakan pada penelitian ini yaitu data sekunder dari SUSENAS (*Survei Sosial Ekonomi Nasional*) tahun 2011 dengan jumlah data 11264 lansia. Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa berdasarkan status lansia bekerja dan tidak bekerja sebanyak 4705 lansia diklasifikasikan bekerja dan 6559 lansia tidak bekerja. Klasifikasi dengan CART menghasilkan bahwa variabel yang berpengaruh adalah hubungan dengan kepala RT, umur, jenis kelamin, lamanya gangguan kesehatan dan ijazah tertinggi.*

Kata Kunci: Klasifikasi, CART, Partisipasi Kegiatan Ekonomi Lansia

***Classification Of Participation Economic Activities
Elderly In East Java Using CART (Classification And
Regression Trees)***

Name : Pitri Ariska Susilowati
NRP : 1313 105 015
Department : Statistika FMIPA-ITS
Supervisor : Dr. Wahyu Wibowo, S.Si, M.Si

Abstract

Economic activity elderly is elderly who do activities with the aim to benefit, at least one hour during the week. Elderly are the people aged 60 years and older. Economic activities are divided into the status of the elderly working and not working. East Java is one of the provinces who have a high percentage of elderly work. The problem is how the classification of the elderly in East Java that works and does not work based on the factors that affect and determine the factors that economic activity participation elderly. CART (Classification And Regression Trees) methods are used in the classification. The data used in this research is secondary data from SUSENAS (National Social Economic Survey) in 2011 with the amount of data 11264 elderly. Based on the analysis has been done, it can be concluded who 4705 are classified work and 6559 classified not working. Factor affecting economic activity elderly in East Java using CART method are household status, age, gender, duration of illness (days) and education level.

Keyword: *Classification, CART, Participation Economic Activities Elderly*

KATA PENGANTAR

Assalamu'alaikum Warahmatullah Wabarokatuh.

Puji syukur alhamdulillah senantiasa penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat, hidayah dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul

“PENGKLASIFIKASIAN PARTISIPASI KEGIATAN EKONOMI LANSIA DI JAWA TIMUR DENGAN MENGGUNAKAN METODE CART (*CLASSIFICATION AND REGRESSION TREES*)”.

Sholawat dan salam tak lupa penulis sampaikan pada junjungan besar Nabi Muhammad SAW. Dalam menyelesaikan laporan Tugas Akhir ini penulis telah banyak menerima bantuan dan dukungan dari berbagai pihak. Oleh karena ucapan terimakasih penulis haturkan kepada :

1. Kedua orang tua tercinta beserta keluarga besar yang telah melimpahkan kasih sayang, segala doa, semangat, dukungan, perhatian yang tiada hentinya kepada penulis.
2. Bapak Dr. Wahyu Wibowo, S.Si, M.Si selaku dosen pembimbing yang telah membimbing saya, memberikan segala masukan, waktu serta pengetahuan demi terselesaikannya Tugas Akhir ini.
3. Ibu Dra. Madu Ratna, S.Si, M.Si dan Ibu Erma Oktania Permatasari, S.Si, M.Si selaku dosen penguji yang telah memberikan banyak saran, kritik dan masukan demi kesempurnaan Tugas Akhir ini.
4. Ibu Dr. Irhamah M,Si selaku dosen wali yang telah memberikan semangat dan memberikan bantuan kepada penulis selama kuliah di Statistika ITS.
5. Bapak Dr. Muhammad Mashuri, MT selaku Ketua Jurusan Statistika ITS dan Ibu Dra. Lucia Aridinanti, MT selaku Ketua Prodi S1 Statistika yang telah memfasilitasi penulis selama menuntut ilmu di Jurusan Statistika ITS.

6. Sahabat-sahabat terimakasih atas dukungan, kebersamaan dan kebahagiaan selama ini. Yaumil, Holis, Ayuk, Lia, Fifi, Anin, Ella terimakasih atas segala motivasi, bantuan dan semangatnya.
7. Teman-teman Lintas Jalur Statitika 2013 seperjuangan dalam penyelesaian Tugas Akhir untuk mencapai Wisuda ke-112 yang saling memberikan motivasi dalam penyelesaian Tugas Akhir.
8. Pihak-pihak yang telah banyak membantu penulis dalam penyusunan Tugas Akhir ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna, oleh karena itu kritik dan saran yang bersifat membangun sangat diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat baik bagi penulis, pembaca, dan semua pihak.

Wassalamu'alaikum Warahmatullah Wabarokatuh.

Surabaya, Juli 2015

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	iii
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvii
BAB I. PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	4
1.4 Manfaat	4
1.5 Batasan Masalah	4
BAB II. TINJAUAN PUSTAKA	
2.1 Penduduk Lanjut Usia	5
2.2 Partisipasi Kegiatan Ekonomi	5
2.3 Statistika Deskriptif.....	7
2.4 <i>Classification And Regression Trees (CART)</i>	7
2.4.1 Pembentukan Pohon Klasifikasi	10
2.4.2 Pemangkasan Pohon Klasifikasi	13
2.4.3 Penentuan Pohon Klasifikasi Optimum.....	14
2.5 Penentuan Skor Variabel <i>Importance</i>	15
2.6 Ukuran Ketepatan Klasifikasi	15
BAB III. METODOLOGI PENELITIAN	
3.1 Sumber Data.....	17
3.2 Variabel Penelitian dan Definisi Operasional	17
3.2.1 Variabel Respon.....	17
3.2.2 Variabel Prediktor	17

3.3 Metode Analisis Data	19
BAB IV. ANALISIS DAN PEMBAHASAN	
4.1 Deskripsi Karakteristik Partisipasi Kegiatan Ekonomi Lansia di Provinsi Jawa Timur	23
4.2 Analisis CART untuk Klasifikasi Partisipasi Kegiatan Ekonomi Lansia di Provinsi Jawa Timur	31
4.2.1 Penentuan untuk Kombinasi Data Learning dan Data Testing	31
4.2.2 Pembentukan Pohon Klasifikasi Maksimal	32
4.2.3 Pemangkasan Pohon Klasifikasi Maksimal	35
4.2.4 Pembentukan Pohon Klasifikasi Optimal	36
4.2.5 Hasil Ketepatan Klasifikasi CART	41
BAB V. KESIMPULAN DAN SARAN	
5.1 Kesimpulan.....	45
5.2 Saran.....	47
DAFTAR PUSTAKA	49
LAMPIRAN	51

DAFTAR TABEL

Tabel 2.1 Tabel Kontingensi $r \times c$	7
Tabel 2.2 Struktur Data Ketepatan Klasifikasi.....	15
Tabel 4.1 Karakteristik Usia Lansia & lamanya kesehatan terganggu	30
Tabel 4.2 Perbandingan Ketepatan Klasifikasi Antar Kombinasi Data	32
Tabel 4.3 Perhitungan Kemungkinan Pemilahan dari Setiap Variabel.....	32
Tabel 4.4 Skor Variabel Terpenting pada Pohon Klasifikasi Maksimal	34
Tabel 4.5 Urutan Pembentukan Pohon Klasifikasi	36
Tabel 4.6 Skor Variabel Terpenting pada Pembentukan Pohon Klasifikasi Optimal.....	38
Tabel 4.7 Klasifikasi Pohon Optimal	41
Tabel 4.8 Validasi Hasil Klasifikasi.....	42

DAFTAR GAMBAR

Gambar 2.1	Struktur Pohon Klasifikasi.....	9
Gambar 3.1	<i>Flowchart</i> Langkah Analisis Penelitian.....	21
Gambar 4.1	Persentase Antara Lansia Bekerja	23
Gambar 4.2	Persentase lansia bekerja& tidak bekerja berdasarkan jenis kelamin	24
Gambar 4.3	Persentase Bidang Pekerjaan Lansia di Provinsi Jawa Timur.....	25
Gambar 4.4	Karakteristik Status/Kedudukan Lansia	25
Gambar 4.5	Karakteristik Status RT Lansia yang Bekerja & Tidak Bekerja.....	26
Gambar 4.6	Karakteristik Ijazah Tertinggi Lansia yang Bekerja & Tidak Bekerja.....	27
Gambar 4.7	Karakteristik Status Perkawinan Lansia yang Bekerja & Tidak Bekerja	28
Gambar 4.8	Karakteristik Klasifikasi Perkotaan&Pedesaan dengan Status Lansia Bekerja&Tidak.....	29
Gambar 4.9	Karakteristik Usia Lansia&Lamanya Kesehatan Terganggu.....	30
Gambar 4.10	Pohon Klasifikasi Maksimal.....	34
Gambar 4.11	Plot <i>Relative Cost</i> dan Jumlah <i>Terminal Nodes</i> dalam Klasifikasi Partisipasi Ekonomi.....	35
Gambar 4.12	Pohon Klasifikasi Optimal.....	37

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Saat ini penduduk Indonesia mengalami struktur penduduk yang menuju proses penuaan dengan ditandainya persentase penduduk lanjut usia (lansia) yang meningkat. Lansia atau usia tua adalah suatu periode penutup dalam rentang hidup seseorang, yaitu suatu periode dimana seseorang telah beranjak jauh dari periode terdahulu yang telah menyenangkan atau beranjak dari waktu yang penuh manfaat. Lansia merupakan seseorang yang berusia 60 tahun atau lebih (hurlock,2000). Menurut Wahyuningsih (2011), diperoleh informasi bahwa dari data Badan Pusat Statistik menunjukkan bahwa penduduk lanjut usia di Indonesia pada tahun 2000 sebanyak 14.439.967 jiwa atau sebesar 7,18 persen, selanjutnya pada tahun 2010 meningkat menjadi 23.992.553 jiwa atau sebesar 9,77 persen. Pada tahun 2020 diprediksikan jumlah lansia mencapai 28.822.879 jiwa atau sekitar 11,34 persen (Wahyuningsih, 2011). Dengan meningkatnya jumlah penduduk lansia ini merupakan cerminan keberhasilan pembangunan terutama pada bidang kesehatan karena semakin meningkat pula usia harapan hidup penduduk di Indonesia.

Menurut data Badan Pusat Statistik (2011), Provinsi Jawa Timur pada tahun 2011 memiliki peringkat kedua setelah Provinsi DI Yogyakarta yang mempunyai penduduk lansia dengan proporsi paling tinggi yaitu memiliki persentase sebesar 10,40 persen. Oleh sebab itu, Provinsi Jawa Timur telah masuk sebagai Provinsi yang berstruktur penduduk tua sebagaimana ketentuan badan dunia, karena jumlah penduduk lanjut usia telah mencapai lebih dari 7 persen. Pertambahan penduduk lansia secara bermakna akan disertai oleh berbagai masalah yang akan yang akan berpengaruh terhadap aspek kehidupan antara lain biologis, fisologis, fisik, mental dan sosial ekonomi (Depkes RI, 1992). Masalah yang terjadi pada lansia yaitu terkait dengan kesehatannya, baik jasmani maupun rohani. Dari kesehatan jasmani terlihat dengan

kemunduran fisiknya yang terjadi secara alamiah. Namun, bukan berarti dalam bidang selain kesehatan seperti kegiatan partisipasi kegiatan ekonomi lansia menjadi permasalahan yang bukan utama. Partisipasi kegiatan ekonomi lansia merupakan kegiatan yang harus mendapatkan penanganan, karena menyangkut pemenuhan kebutuhan hidup manusia. Partisipasi kegiatan ekonomi yang terjadi pada lansia ini dilihat dari keikutsertaan lansia tersebut dalam hal bekerja atau tidak bekerja. Berdasarkan data hasil Sakernas 2011, penduduk lansia Jawa Timur masih banyak yang tergolong sebagai lansia produktif. Dari jumlah keseluruhan penduduk lansia sekitar 45,4 persen diantaranya masih bekerja, sedangkan sisanya pengangguran sebesar 54,6 persen. Tingginya partisipasi penduduk lansia yang bekerja antara lain untuk memenuhi kebutuhan ekonomi rumah tangga, mengisi waktu luang dan untuk menjaga kesehatan badan.

Karena tingginya persentase lansia yang bekerja maka pada penelitian ini kali ini akan membahas tentang klasifikasi partisipasi kegiatan ekonomi lansia berdasarkan faktor yang mempengaruhinya di Provinsi Jawa Timur dengan menggunakan metode CART (*Classification and Regression Trees*). Kelebihan metode CART (*Classification and Regression Trees*) ini yaitu bersifat nonparametrik sehingga tidak memerlukan asumsi-asumsi yang mengikat seperti asumsi distribusi normal untuk variabel prediktor, struktur data dapat dilihat secara visual sehingga memudahkan eksplorasi dan pengambilan keputusan, tidak hanya memberikan klasifikasi namun juga estimasi probabilitas kesalahan pengklasifikasian, kemudahan dalam menginterpretasikan hasil (Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J., 1993).

Penelitian tentang lansia sebelumnya pernah dilakukan oleh Cristie (2012) meneliti tentang faktor-faktor yang berpengaruh terhadap aktivitas ekonomi lansia di Provinsi Jawa Timur dengan menggunakan Regresi Logistik Biner. Hasil penelitian tersebut menyebutkan bahwa terdapat faktor yang berpengaruh pada aktivitas ekonomi lansia yaitu umur, jenis kelamin, hubungan

dengan kepala rumah tangga, status perkawinan, ijazah tertinggi, dapat membaca dan menulis huruf latin maupun huruf arab. Penelitian penggunaan metode CART (*Classification and Regression Trees*) pernah dilakukan oleh Pratiwi (2014) tentang klasifikasi pengangguran terbuka menggunakan metode CART di Provinsi Sulawesi Utara. Hasil penelitian menyimpulkan bahwa faktor yang mempengaruhi angkatan kerja Provinsi Sulawesi Utara untuk bekerja atau tidak yaitu status dalam rumah tangga, jenis kelamin, usia, pendidikan terakhir dan status perkawinan. Penelitian lain terkait metode CART pernah dilakukan oleh Andriyashin (2005) yang menyatakan bahwa metode CART bukan saja merupakan alat statistik nonparametrik yang efektif untuk kasus klasifikasi/regresi tetapi juga cukup kuat dalam analisis eksplorasi data. Menelaah dari penelitian terkait lansia dan metode CART, maka dalam penelitian ini digunakan metode CART untuk mendapatkan klasifikasi partisipasi kegiatan ekonomi lansia di Provinsi Jawa Timur berdasarkan faktor yang mempengaruhinya.

Penelitian ini menggunakan data sekunder yang diperoleh dari data hasil SUSENAS (Survei Sosial Ekonomi Nasional) tahun 2011 Provinsi Jawa Timur. Pengklasifikasian partisipasi kegiatan ekonomi lansia berdasarkan faktor-faktor yang mempengaruhinya dapat dilakukan dengan beberapa metode, salah satunya yaitu menggunakan metode CART (*Classification and Regression Trees*). Metode CART (*Classification and Regression Trees*) bertujuan untuk mengklasifikasi suatu kelompok observasi kedalam suatu kelompok dari suatu kelas-kelas yang diketahui. Metode CART (*Classification and Regression Trees*) dapat diterapkan untuk data dengan jumlah besar, variabel yang sangat banyak dan melalui prosedur pemilihan biner (Lewis, 2000).

2.2 Rumusan Masalah

Pertambahan penduduk lansia secara bermakna akan disertai oleh berbagai masalah salah satunya masalah ekonomi karena berhubungan dengan pemenuhan kebutuhan hidup manusia Masalah ekonomi pada lansia dilihat dari partisipasi kegiatan

ekonomi lansia yang dibedakan menjadi status lansia bekerja dan status lansia tidak bekerja. Terdapat beberapa faktor yang berpengaruh terhadap partisipasi kegiatan ekonomi lansia yaitu jenis kelamin, umur, status dalam rumah tangga, ijazah tertinggi lansia, status perkawinan, klasifikasi desa/kelurahan tempat tinggal, dapat membaca dan menulis huruf latin atau tidak dan lamanya lansia mengalami gangguan kesehatan. Provinsi Jawa Timur merupakan salah satu provinsi yang memiliki persentase lansia bekerja yang tinggi. Permasalahannya adalah bagaimana klasifikasi lansia di Provinsi Jawa Timur yang bekerja dan tidak bekerja berdasarkan faktor-faktor yang mempengaruhinya dan mengetahui faktor-faktor yang berpengaruh terhadap partisipasi kegiatan ekonomi lansia.

2.3 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah

1. Memperoleh urutan variabel yang mempengaruhi partisipasi kegiatan ekonomi lansia di Provinsi Jawa Timur.
2. Memperoleh model pohon klasifikasi dan hasil ketepatan klasifikasi partisipasi kegiatan ekonomi lansia di Provinsi Jawa Timur.

2.4 Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan bahan masukan kepada pemerintah provinsi Jawa Timur mengenai klasifikasi partisipasi kerja lansia berdasarkan faktor yang mempengaruhinya sehingga pemerintah dapat membuat kebijakan dalam mengatasi masalah lansia, terutama masalah tenaga kerja lansia di Provinsi Jawa Timur.

2.5 Batasan Penelitian

Batasan masalah dalam penelitian ini yaitu variabel yang digunakan hanya variabel yang terdapat dalam SUSENAS 2011 dan responden SUSENAS pada penelitian ini adalah individu yang berusia 60 tahun keatas.

BAB II

TINJAUAN PUSTAKA

2.1 Penduduk Lanjut Usia

Menurut Badan Pusat Statistik, penduduk adalah semua orang yang berdomisili di wilayah geografis Republik Indonesia selama 6 bulan atau lebih dan atau mereka yang berdomisili kurang dari 6 bulan tetapi bertujuan untuk menetap. Penduduk lanjut usia (lanjut usia) yaitu penduduk yang apabila telah berumur 60 tahun ke atas untuk negara berkembang atau 65 tahun ke atas untuk negara maju. Di antara lansia yang berusia 60 tahun ke atas dikelompokkan lagi menjadi *young old* (60-69 tahun), *old* (70-79 tahun) dan *old-old* (80 tahun ke atas). Dari aspek kesehatan, seseorang disebut sebagai lansia jika berusia 60 tahun ke atas. Sedangkan penduduk yang berusia antara 49-69 tahun disebut sebagai *pra-sensile*. Lansia yang berusia 70 tahun ke atas disebut sebagai lansia beresiko. Dari aspek ekonomi, lansia (60 tahun ke atas) dikelompokkan menjadi lansia yang produktif yaitu lansia yang sehat baik dari aspek fisik, mental maupun sosial dan lansia yang tidak produktif yaitu lansia yang sehat secara fisik, tetapi tidak sehat dari aspek mental dan sosial atau sehat secara mental tetapi tidak sehat dari aspek fisik dan sosial atau lansia yang tidak sehat baik dari aspek fisik, mental maupun sosial.

2.2 Partisipasi Kegiatan Ekonomi

Menurut Badan Pusat Statistik, definisi kegiatan ekonomi yaitu melakukan kegiatan dengan tujuan mendapat pendapatan atau keuntungan, minimal satu jam selama seminggu, sedangkan yang bukan melakukan kegiatan ekonomi adalah orang-orang yang mengurus rumah tangga, sekolah atau melakukan kegiatan lainnya seperti pensiun.

Menurut Badan Pusat Statistik, angkatan kerja lansia merupakan penduduk lansia yang bekerja dan mencari pekerjaan (pengangguran). Penduduk lansia ini sering disebut penduduk lansia potensial. Mereka tergolong sebagai penduduk lansia yang produktif dan mandiri. Penduduk lansia potensial banyak ditemukan di negara berkembang

dan negara-negara yang belum memiliki tunjangan sosial untuk hari tua. Mereka berusaha tetap bekerja dalam upaya memenuhi tuntutan hidup maupun mencukupi kebutuhan keluarga yang menjadi tanggungannya. Usaha pemberdayaan penduduk lansia merupakan salah satu usaha melatih kemandirian. Bidang pekerjaan utama lansia di Provinsi Jawa Timur terbagi menjadi 3 sektor yaitu sektor pertanian, sektor jasa dan sektor industri. Sektor pertanian meliputi pertanian tanaman padi, perkebunan, hortikultura, perikanan dan kehutanan. Sektor jasa meliputi jasa pendidikan, jasa kesehatan, jasa kemasyarakatan, informasi dan komunikasi, keuangan dan asuransi, konstruksi dan bangunan. Sektor industri meliputi industri pengolahan, pertambangan dan penggalian, listrik dan gas [4]. Status atau kedudukan lansia dalam pekerjaan utama terbagi menjadi formal dan non formal. Kedudukan formal seperti pekerjaan sebagai karyawan atau pegawai, sedangkan kedudukan non formal meliputi berusaha sendiri, pekerja bebas, pekerja keluarga.

Undang-Undang nomor 13 tahun 1998 pasal 1 ayat 3 menjelaskan bahwa lanjut usia potensial adalah lanjut usia yang masih mampu melakukan pekerjaan dan atau kegiatan yang dapat menghasilkan barang atau jasa. Selanjutnya pasal 1 ayat 4 menjelaskan bahwa lanjut usia tidak potensial adalah lanjut usia yang tidak berdaya mencari nafkah sehingga hidupnya bergantung pada bantuan orang lain. Oleh karena itu pembangunan bidang ketenagakerjaan tidak hanya ditujukan pada penduduk muda yang produktif, tetapi juga diarahkan kepada penduduk lansia potensial. Hak untuk mendapatkan pelayanan kesempatan kerja terdapat dalam Undang-Undang nomor 13 tahun 1998 bab 3 pasal 5 ayat 2c. Pelayanan kesempatan kerja bagi lanjut usia potensial dimaksudkan memberi peluang untuk mendayagunakan pengetahuan, keahlian, kemampuan, keterampilan dan pengalaman yang dimilikinya pada sektor formal dan nonformal, melalui perseorangan, kelompok atau organisasi atau lembaga baik pemerintah maupun masyarakat tercantum pada Undang-Undang nomor 13 tahun 1998 tentang kesejahteraan lansia bab 6 pasal 15 ayat 1 dan 2.

2.3 Statistika Deskriptif

Statistika deskriptif merupakan serangkaian teknik statistika yang digunakan untuk mempelajari cara-cara pengumpulan, penyusunan, penyajian data, dan penarikan kesimpulan secara numerik (Kurniawan, 2010). Statistika deskriptif memberikan informasi hanya mengenai data yang akan diteliti dan sama sekali tidak menarik kesimpulan apapun tentang gugus data induknya yang lebih besar. Ada 2 macam penyajian data yang sering dipakai, yaitu tabel atau daftar dan grafik atau diagram (Walpole, 1995).

Tabel kontingensi atau yang sering disebut tabulasi (*cross tabulation atau cross classification*) adalah tabel yang berisi data jumlah atau frekuensi atau beberapa klasifikasi (kategori). *Cross tabulation* yaitu suatu metode statistik yang menggambarkan dua atau lebih variabel secara simultan dan hasilnya ditampilkan dalam bentuk tabel yang merefleksikan distribusi bersama dua atau lebih variabel dengan jumlah kategori yang terbatas (Agresti, 2007). Tabel kontingensi adalah sebagai berikut.

Tabel 2.1 Tabel Kontingensi $r \times c$

Baris	Kolom			
	1	2	..	C
1	n_{11}	n_{12}	..	n_{1c}
2	n_{21}	n_{22}	..	n_{2c}
:	:	:	:	:
R	n_{r1}	n_{r2}	..	n_{rc}

2.4 Classification And Regression Trees (CART)

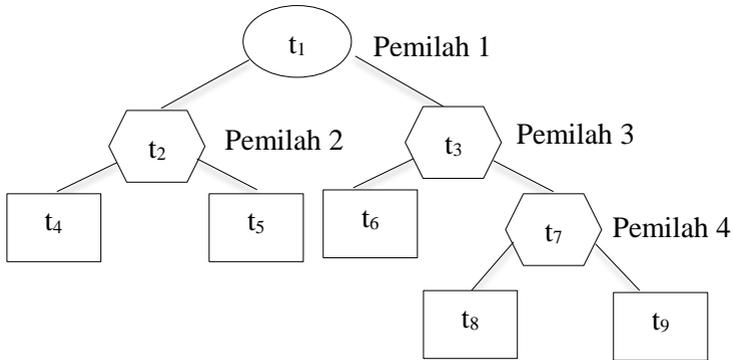
CART merupakan salah satu metode nonparametrik yang digunakan untuk teknik pohon keputusan. Metode ini digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Apabila variabel responnya berbentuk kontinyu maka CART yang dihasilkan adalah regresi pohon, namun jika variabel responnya berbentuk katagorik maka yang yang dihasilkan adalah klasifikasi pohon (Breiman dkk, 1993). Penggunaan variabel dalam suatu pengklasifikasian menurut Roger dan Lewis

(2000) ada 4 komponen, yaitu variabel respon, variabel prediktor, data *learning* dan data *testing*. Variabel respon merupakan satu karakteristik yang diharapkan dapat diprediksi dengan menggunakan variabel prediktor. Variabel prediktor merupakan suatu variabel yang dihubungkan dengan variabel yang akan diprediksi. Sehingga secara umum ada banyak variabel prediktor yang mungkin mempengaruhi variabel respon. Himpunan data *learning*, yaitu suatu himpunan data yang terdiri dari nilai-nilai variabel prediktor dan variabel respon yang berasal dari sekumpulan kasus. Kasus-kasus yang terdapat dalam data *learning* mempunyai kemiripan dengan kasus yang akan diprediksi hasilnya di masa depan. Sedangkan untuk data *testing*, yaitu himpunan data dari kasus-kasus yang akan diprediksi hasilnya secara tepat. Dapat juga dikatakan data ini digunakan untuk mengetahui seberapa tepat model yang sudah dibentuk atau diklasifikasi yang dihasilkan oleh data *learning*.

Roger dan Lewis (2000) menyatakan bahwa pohon keputusan dengan menggunakan algoritma penyeleksian rekursif secara biner (*binary recursive partitioning*). Istilah “*binary*” diartikan sebagai pemilahan yang dilakukan pada sekelompok data yang terkumpul dalam suatu ruang yang disebut simpul atau *node* menjadi dua kelompok yang disebut sebagai simpul anak (*child nodes*). “*Recursive*” menunjukkan bahwa prosedur penyekatan secara biner dilakukan secara berulang-ulang. Penyekatan *recursive* diperoleh dari penyekatan simpul awal kemudian dapat dipilah kembali menjadi dua simpul anak lagi, begitu seterusnya dan berhenti jika memenuhi kriteria tertentu. Sedangkan istilah “*partitioning*” artinya bahwa proses klasifikasi dilakukan dengan cara memilah suatu kumpulan data menjadi beberapa bagian.

Pada pohon klasifikasi simpul asli disebut *parent node* (simpul induk) yang mengandung semua data dengan notasi t_1 , beberapa simpul dalam dan simpul akhir. Pada Gambar 2.1 simpul dalam (*internal nodes*) dilambangkan dengan segi enam yaitu t_2 , t_3 , t_7 , sedangkan simpul akhir (*terminal node*) dengan segi empat yaitu t_4 , t_5 , t_6 , t_8 , dan t_9 . Untuk menghitung kedalaman (depth) dimulai dari simpul pertama t_1 yang berada pada kedalaman 1,

sedangkan t_2 dan t_3 pada kedalaman 2, begitu seterusnya sampai pada simpul terminal t_8 dan t_9 yang berada dalam kedalaman 4. Dengan cara ini, akan diperoleh sekatan-sekatan dengan perbedaan yang sangat besar dengan respon yang lebih homogen dalam tiap-tiap sekatan. Ilustrasi struktur pohon klasifikasi terdapat pada Gambar 2.1 berikut



Gambar 2.1 Struktur Pohon Klasifikasi

Adapun beberapa kelebihan metode klasifikasi pohon ini antara lain.

1. Metode ini bersifat nonparametrik sehingga tidak memerlukan asumsi-asumsi yang mengikat seperti asumsi distribusi normal untuk variabel prediktor.
2. Struktur data dapat dilihat secara visual sehingga memudahkan eksplorasi dan pengambilan keputusan berdasarkan model yang diperoleh.
3. Tidak hanya memberikan klasifikasi, namun juga estimasi probabilitas kesalahan pengklasifikasian
4. Mampu mengidentifikasi interaksi antar variabel prediktor yang berpengaruh secara lokal akibat diterapkannya pengambilan keputusan secara bertahap dalam himpunan-himpunan bagian data pengukuran yang kompleks.
5. Hasil klasifikasi akhir berbentuk sederhana mengklasifikasikan data baru secara efisien
6. Kemudahan dalam interpretasi hasil

Langkah-langkah penerapan algoritma CART adalah sebagai berikut.

1. Pembentukan Pohon Klasifikasi
2. Pemangkasan Pohon Klasifikasi
3. Penentuan Pohon Klasifikasi Optimal

2.4.1 Pembentukan Pohon Klasifikasi

Untuk membentuk pohon klasifikasi pertama ditentukan dahulu variabel apa dan nilai mana dari variabel tersebut yang layak dijadikan pemilah untuk tiap simpul. Pembentukan pohon klasifikasi dibutuhkan *learning* sampel L yang terdiri atas N amatan. Ukuran pohon akan menentukan kekuatan prediktif/akurasi pohon klasifikasi. Proses pembentukan pohon klasifikasi terdiri dari tiga tahapan yaitu seperti berikut.

A. Pemilihan Pemilah

Untuk membentuk pohon klasifikasi digunakan sampel data *learning* (L) yang masih bersifat heterogen. Sampel tersebut akan dipilih berdasarkan aturan pemilahan dan kriteria *goodness of split*. Pemilihan pemilah tergantung pada jenis pohon atau lebih tepatnya pada jenis variabel respon. Himpunan bagian yang dihasilkan dari pemilahan harus lebih homogen dibandingkan dengan pemilahan sebelumnya. Langkah yang harus dilakukan sebagai berikut.

- i. Mendefinisikan fungsi keheterogenan simpul

Fungsi keheterogenan yang sangat mudah dan sesuai diterapkan dalam berbagai kasus adalah Indeks Gini (Breiman dkk, 1993). Indeks Gini akan selalu memisahkan kelas yang anggota kelasnya terbesar lebih dahulu atau yang merupakan kelas terpenting dalam simpul tersebut. Pemilah yang memberikan nilai penurunan keheterogenan tertinggi adalah pemilah terbaik. Berikut merupakan fungsi keheterogenan Indeks Gini.

$$i(t) = \sum_{j \neq 1} p(i|t)p(j|t) \quad (2.1)$$

dengan,

$i(t)$: Fungsi keheterogenan indeks gini pada simpul t

$p(i|t)$: Proporsi kelas i pada simpul t

$p(j|t)$: Proporsi kelas j pada simpul t

ii. Pemilahan Simpul

Pemilahan dimulai dengan memeriksa nilai-nilai variabel prediktor dan dilakukan secara rekursif pada tiap simpul dengan dua tahap. Pertama adalah mencari semua kemungkinan pemilah pada variabel prediktor. Menurut Breiman (1993), aturan pemilah setiap simpul menjadi 2 simpul anak adalah sebagai berikut.

1. Setiap pemilahan hanya bergantung pada nilai yang berasal dan satu variabel prediktor.
2. Untuk variabel kontinyu, pemilahan ditentukan oleh pertanyaan “Apakah $X_j \leq C_t$ ”. Dimana $j = 1, 2, \dots, n-1$, sedangkan C_t adalah nilai tengah dari dua nilai amatan variabel X_j , berurutan yang berbeda. Jika ruang sampel berukuran n dan terdapat n nilai amatan yang berbeda pada variabel X_j , maka akan terdapat sebanyak-banyaknya $n-1$ pemilahan yang berbeda.
3. Untuk variabel kategori, pemilahan berasal dari semua kemungkinan pemilahan berdasarkan terbentuknya 2 simpul yang saling lepas (*disjoint*). Jika variabel X_j adalah kategori nominal bertaraf L , maka akan diperoleh $2^{L-1}-1$ pemilahan yang mungkin. Tetapi jika variabel X_j adalah kategori ordinal maka akan diperoleh $L-1$ pemilahan yang mungkin.

Langkah selanjutnya adalah mencari pemilah terbaik dari setiap variabel prediktor. Pemilah yang memaksimumkan ukuran kehomogenan masing-masing simpul anak relatif terhadap simpul induknya dan memaksimumkan ukuran pemisahan antara 2 simpul anak tersebut adalah pemilah terbaik. Pemilah yang terpilih akan membentuk himpunan kelas yang disebut simpul satu. Setiap pemilahan akan dilakukan pada setiap simpul sampai diperoleh simpul akhir dan menghasilkan 2 kelas yaitu simpul anak. Jika terdapat data *missing value* (pengamatan yang hilang) pada variabel pemilah, maka pemilahan akan digantikan variabel lain yang disebut *surrogate*

splitter (pemilah pengganti), dimana nilai asosiasinya terbesar diantara pemilah pengganti lain.

iii. Menentukan kriteria *goodness of split*

Goodness of split merupakan suatu evaluasi bagi pemilahan yang dilakukan oleh pemilah s pada simpul t . Misal, terdapat pemilah s yang akan memilah t menjadi simpul kiri (t_L) dengan proporsi (p_L) dan simpul kanan (t_R) dengan proporsi (p_R), maka *goodness of split* $\varphi(s, t)$ didefinisikan sebagai penurunan keheterogenan.

$$\varphi(s, t) = \Delta_i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.2)$$

Dimana,

$\varphi(s, t)$: kriteria *goodness of split*

$i(t)$: fungsi keheterogenan pada simpul

p_L : proporsi pengamatan menuju simpul kiri

p_R : proporsi pengamatan menuju simpul kanan

$i(t_L)$: fungsi keheterogenan pada simpul anak kiri

$i(t_R)$: fungsi keheterogenan pada simpul anak kanan

Pengembangan pohon dilakukan dengan pencarian semua pemilah yang mungkin pada simpul t_1 . Simpul t_1 akan dipilah menjadi t_2 dan t_3 dengan pemilah s^* yang memberikan nilai penurunan keheterogenan tertinggi, yaitu

$$\nabla_i(s^*, t_1) = \max_{s \in S} \Delta_i(s, t) \quad (2.3)$$

Cara yang sama dilakukan pada t_2 dan t_3 secara terpisah dan seterusnya.

B. Penentuan Simpul Terminal

Simpul t dapat dijadikan simpul terminal apabila tidak terdapat penurunan keheterogenan yang berarti sehingga tidak akan dipilah lagi, hanya terdapat suatu pengamatan pada tiap simpul anak atau adanya batasan minimum n . Menurut Breiman dkk (1993), pengembangan pohon akan berhenti apabila pada simpul terdapat $n_1 < 5$. Selain itu, proses pembentukan pohon juga dapat berhenti bila sudah mencapai batasan jumlah level yang telah ditentukan atau

tingkat kedalaman dalam pohon maksimal yang kemudian pohon akan berhenti.

C. Penentuan label kelas

Penentuan label kelas pada simpul terminal berdasarkan aturan jumlah terbanyak yaitu jika.

$$p(j_0|t) = \max p(j|t) = \max \frac{N_j(t)}{N_t} \quad (2.4)$$

Dimana,

$p(j t)$: proporsi kelas j pada simpul t
$N_j(t)$: jumlah pengamatan kelas j pada simpul t
N_t	: jumlah pengamatan pada simpul t

Label kelas untuk simpul terminal t adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t yang paling kecil sebesar $r(t) = 1 - \max p_j(j|t)$

2.4.2 Pemangkasan Pohon Klasifikasi

Pohon yang dibentuk dengan aturan pemilah dan kriteria *goodness of split* berukuran sangat besar karena penghentian pohon berdasarkan banyaknya amatan pada simpul terminal atau besarnya tingkat kehomogenan. Semakin banyak pemilahan yang dilakukan mengakibatkan makin kecilnya tingkat kesalahan prediksi atau dapat disebut juga dengan kasus *overfitting* yaitu nilai prediksi melebihi nilai yang sebenarnya. Selain itu, bila dalam proses pemilahan diberikan batasan, padahal pemilahan masih layak untuk dilakukan maka akan terjadi kasus *underfitting*. Oleh karena itu, untuk mendapatkan pohon yang layak maka perlu dilakukan pemangkasan (*pruning*) yaitu suatu penilaian ukuran pohon tanpa mengorbankan ketepatan atau kebaikannya melalui pengurangan simpul pohon sehingga dicapai ukuran pohon yang layak. Untuk mendapatkan ukuran pohon yang layak dilakukan pemangkasan pohon dengan *cost complexity minimum* (Breiman dkk, 1993). Untuk $\alpha \geq 0$ maka ukuran *cost complexity* adalah.

$$R_\alpha(T) = R(T) + \alpha(\bar{T}) \quad (2.5)$$

Dimana,

$R(T)$: *resubstitution estimate* (penduga pengganti)

α : parameter *cost complexity* bagi penambahan satu simpul akhir pada pohon T

$R_\alpha(T)$: *resubstitution* suatu pohon T pada kompleksitas α

$|\bar{T}|$: Banyaknya simpul terminal pohon T

Cost complexity pruning menentukan pohon bagian $T(\alpha)$ yang meminimumkan $R_\alpha(T)$ pada seluruh pohon bagian untuk setiap nilai α . Nilai parameter kompleksitas α akan secara perlahan meningkat selama proses pemangkasan. Selanjutnya pencarian pohon bagian $T(\alpha) < T_{maks}$ yang dapat meminimumkan $R_\alpha(T)$ yaitu:

$$R_\alpha(T(\alpha)) = \min_{T < T_{maks}} R_\alpha(T) \quad (2.6)$$

2.4.3 Penentuan Pohon Klasifikasi Optimum

Pohon klasifikasi yang berukuran sangat besar memberikan nilai penduga yang sangat kecil, sehingga pohon ini lebih dipilih untuk menduga nilai respon. Tetapi ukuran pohon yang besar bersifat *overfitting* sehingga menyebabkan nilai kompleksitas yang tinggi. Karena struktur data yang digambarkan cenderung kompleks, sehingga perlu pemilihan pohon yang optimum dengan ukuran sederhana dan memberikan nilai penduga pengganti cukup kecil.

Penduga pengganti yang sering digunakan apabila amatan yang ada cukup besar adalah *Test Sample Estimate*. Prosedur ini diterapkan dengan membagi sampel L menjadi dua himpunan L_1 dan L_2 . Amatan dalam L_1 digunakan untuk membentuk pohon T. Sedangkan amatan pada L_2 digunakan untuk menduga $R(T)$. Jika N_2 merupakan jumlah amatan dalam L_2 dan $X(\cdot)$ bernilai 0 jika pernyataan dalam tanda kurung salah dan bernilai 1 jika pernyataan dalam tanda kurung benar. Penduga sampel uji adalah sebagai berikut.

$$R^{ts}(T_t) = \frac{1}{N_2} \sum (x_n \cdot j_n) \in L_2 X(d(x_n) \neq j_n) \quad (2.7)$$

Pohon klasifikasi yang optimal adalah pohon T^* yang memiliki nilai penduga sampel uji minimum.

2.5 Penentuan Skor Variabel *Importance*

Penentuan skor pada variabel *importance* langkah-langkahnya sebagai berikut.

1. Menentukan tingkat kedalaman (*depth*) pohon klasifikasi yang terbentuk
2. Menghitung nilai *goodness of split* pada setiap kemungkinan pemilah yang mungkin pada setiap tingkat kedalaman pada pohon klasifikasi
3. Menjumlahkan pemilah yang sama dari tingkat kedalaman (*depth*) yang terbentuk.
4. Selanjutnya nanti akan diperoleh skor (%) pada variabel *importance*.

2.6 Ukuran Ketepatan Klasifikasi

Cara yang digunakan untuk mengevaluasi hasil ketepatan klasifikasi adalah dengan menghitung akurasi klasifikasi. Pada penelitian ini evaluasi akurasi klasifikasi ialah dengan menghitung nilai 1-APER (*apparent error rate*) atau biasa disebut dengan nilai total akurasi. Tingkat akurasi total merupakan proporsi observasi yang diprediksi secara benar oleh fungsi klasifikasi. Nilai total tingkat kesalahan merupakan proporsi observasi yang diprediksi secara tidak benar oleh fungsi klasifikasi (Johnson dan Winchern, 2007). *Sensitivity* menggambarkan akurasi pada sampel kelas i sedangkan *specificity* menggambarkan akurasi pada sampel kelas j . Metode klasifikasi yang baik seharusnya mampu mengukur *sensitivity* dan *specificity* sama baiknya.

Tabel 2.2 Struktur Data Ketepatan Klasifikasi

Kelas Observasi Y	Kelas Prediksi		Total
	1	2	
1	n_{11}	n_{12}	N_1
2	n_{21}	n_{22}	N_2
Total	N_1	N_2	N

Keterangan:

n_{11} : Jumlah observasi dari variabel Y kelas 1 yang tepat diprediksi sebagai variabel Y kelas 1

n_{12} : Jumlah observasi dari variabel Y kelas 1 yang salah diprediksi sebagai variabel Y kelas 2

n_{21} : Jumlah observasi dari variabel Y kelas 2 yang salah diprediksi sebagai variabel Y kelas 1

n_{22} : Jumlah observasi dari variabel Y kelas 2 yang tepat diprediksi sebagai variabel Y kelas 2

N_1 : Jumlah observasi dari variabel Y kelas 1

N_2 : Jumlah observasi dari variabel Y kelas 2

N : Jumlah total observasi/prediksi

$$\text{Sensitivity} = \frac{n_{11}}{N_1} \quad (2.8)$$

$$\text{Specificity} = \frac{n_{22}}{N_2} \quad (2.9)$$

$$\text{Apparent Error Rate (APER)} = \frac{\text{Jumlah prediksi salah}}{\text{Jumlah total prediksi}} = \frac{n_{21}+r}{N} \quad (2.10)$$

BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder. Data yang diperoleh dari hasil Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2011 Provinsi Jawa Timur dengan jumlah 38 kabupaten/kota dan jumlah sampel penduduk yang usianya 60 tahun keatas sebanyak 11264 lansia.

3.2 Variabel Penelitian dan Definisi Operasional

Variabel penelitian yang digunakan dalam penelitian ini yaitu terdiri dari variabel respon dan variabel prediktor. Di bawah ini dijelaskan variabel respon dan variabel prediktor yang dipakai dalam penelitian ini.

3.2.1 Variabel Respon

Variabel respon (Y) yang digunakan dalam penelitian ini yaitu partisipasi kegiatan ekonomi lansia di Jawa Timur yang dilihat dari lansia tersebut bekerja atau tidak dalam 1 bulan terakhir. Variabel respon (Y) tersebut mempunyai skala nominal dengan dibedakan menjadi 2 yaitu.

Y = 1, jika lansia bekerja

Y = 2, jika lansia tidak bekerja

Lansia yang bekerja merupakan lansia yang melakukan kegiatan dengan tujuan mendapatkan upah atau penghasilan atau keuntungan minimal satu jam dalam seminggu. Sedangkan lansia tidak bekerja merupakan mereka mengurus rumah tangga ataupun sedang menempuh sekolah.

3.2.2 Variabel Prediktor

Variabel prediktor yang digunakan dalam penelitian ini yaitu variabel yang diduga berpengaruh terhadap partisipasi kegiatan ekonomi lansia di Provinsi Jawa Timur. Variabel tersebut adalah sebagai berikut.

1. Jenis Kelamin (X_1), dikategorikan menjadi 2 yaitu sebagai berikut.

$X_1 = 1$, jika lansia mempunyai jenis kelamin laki-laki

$X_1 = 2$, jika lansia mempunyai jenis kelamin perempuan

2. Umur (X_2) dihitung dalam tahun, dalam penelitian ini umur memiliki skala rasio

3. Hubungan dengan kepala rumah tangga (X_3) yaitu hubungan lansia dengan kepala rumah tangga. Variabel ini dibedakan menjadi 4 kategori yaitu sebagai berikut.

$X_3 = 1$, jika lansia sebagai kepala rumah tangga

$X_3 = 2$, jika lansia sebagai istri

$X_3 = 3$, jika lansia sebagai orang tua atau mertua

$X_3 = 4$, jika lansia sebagai anggota rumah tangga lainnya

4. Pendidikan terakhir atau ijazah tertinggi yang dimiliki (X_4) yaitu merupakan pendidikan terakhir yang ditempuh lansia atau ijazah tertinggi oleh lansia sekarang. Variabel ini dibedakan menjadi 5 kategori sebagai berikut.

$X_4 = 1$, jika lansia tidak memiliki ijazah/tidak tamat SD

$X_4 = 2$, jika lansia memiliki ijazah tertinggi SD

$X_4 = 3$, jika lansia memiliki ijazah tertinggi SMP/Sederajat

$X_4 = 4$, jika lansia memiliki ijazah tertinggi SMA/Sederajat

$X_4 = 5$, jika lansia memiliki ijazah tertinggi perguruan tinggi

5. Status perkawinan (X_5) merupakan status perkawinan yang dimiliki oleh lansia pada saat dilakukan survei. Lansia yang disebut kawin adalah lansia yang mempunyai suami atau istri yang masih hidup saat dilakukan survei. Lansia yang bercerai adalah lansia yang berpisah dengan suami atau istri yang disahkan secara hukum atau agama dan belum kawin lagi. Lansia yang ditinggal meninggal oleh suami atau istrinya dan belum kawin lagi masuk dalam kriteria lansia bercerai atau disebut juga dengan cerai mati. Status ini dikategorikan menjadi 3 sebagai berikut.

$X_5 = 1$, jika lansia berstatus belum kawin

$X_5 = 2$, jika lansia berstatus kawin

$X_5 = 3$, jika lansia berstatus cerai (cerai-hidup/cerai-mati)

6. Klasifikasi desa/kelurahan tempat tinggal (X_6) yaitu lansia yang bertempat tinggal dengan dibedakan yaitu perkotaan atau pedesaan dapat dikategorikan sebagai berikut.
 $X_6 = 1$, jika lansia bertempat tinggal di daerah perkotaan
 $X_6 = 2$, jika lansia bertempat tinggal di daerah pedesaan
7. Dapat membaca dan menulis huruf latin (X_7) atau dapat disebut juga lansia buta huruf latin atau tidak. Buta huruf latin dapat dikategorikan menjadi.
 $X_7 = 1$, jika lansia dapat membaca dan menulis huruf latin
 $X_7 = 2$, jika lansia tidak dapat membaca dan menulis huruf latin
8. Keluhan kesehatan yang dirasakan oleh penduduk lansia dalam satu bulan terakhir. Indikator keterangan kesehatan yang terdapat di kuesioner berbagai macam jenis keluhan kesehatan. Keluhan kesehatan yang dimaksud antara lain panas, batuk, pilek, diare, sakit kepala berulang, sakit gigi serta penyakit lainnya. Keluhan kesehatan lansia ini dilihat dari lamanya terganggu. Lamanya terganggu ini memiliki skala rasio.

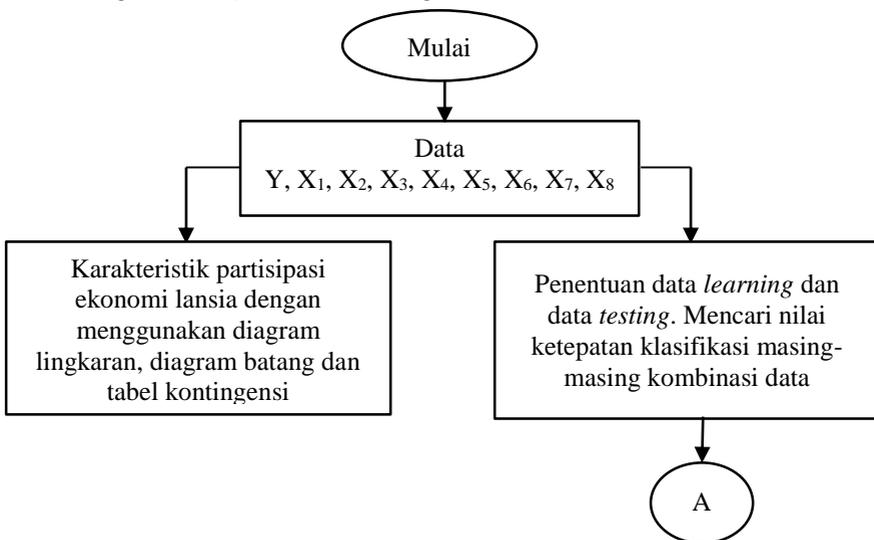
3.3 Metode Analisis Data

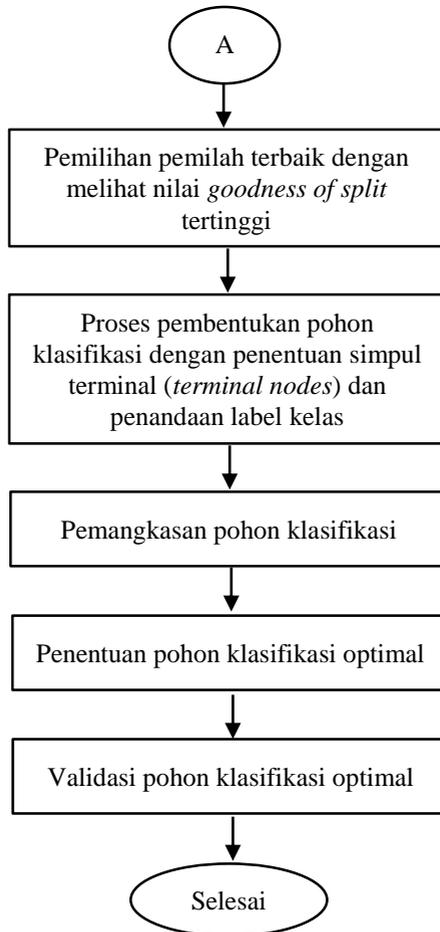
Langkah yang dilakukan dalam penelitian ini, pertama yang dilakukan identifikasi karakteristik dari data SUSENAS berdasarkan variabel-variabel yang digunakan. Selanjutnya dilakukan analisis CART untuk melihat klasifikasi partisipasi kegiatan ekonomi lansia berdasarkan faktor yang mempengaruhinya. Berikut langkah-langkah analisis CART.

1. Membagi data menjadi beberapa kombinasi proporsi data *learning* dan data *testing*. Kombinasi yang digunakan data *learning* dan data *testing* yaitu (95% : 5%), (90% : 10%), (85% : 15%), (80% : 20%) dan (75% : 25%). Kemudian mendapatkan ketepatan klasifikasi terbesar antara beberapa kombinasi data *learning* dan *testing* tersebut.
2. Menentukan pemilah terbaik dari semua variabel prediktor untuk memilah secara rekursif biner pada setiap simpul yang akan dipilah, yaitu dengan kriteria pemilah yang memberikan penurunan keheterogenan tinggi.

3. Proses pemilahan akan terus dilakukan pada simpul (*node*) yang masih bisa dipilah menjadi simpul anak sampai tidak bisa dipilah kembali yang nantinya disebut sebagai *terminal node*.
4. Penandaan label kelas, label kelas dari *terminal node* ditentukan berdasarkan aturan jumlah terbanyak dari kelas yang ada pada variabel respon.
5. Pemangkasan pohon klasifikasi, penghentian pembentukan pohon klasifikasi dilakukan dengan menggunakan kriteria ukuran *cost complexity minimum*.
6. Pemilihan pohon klasifikasi optimal melalui *test sample estimates*. Pemilihan model pohon klasifikasi terbaik, menggunakan kriteria *test relative cost*. Setelah didapatkan nilai *test relative cost* paling *minimum* yang nantinya akan digunakan sebagai penentuan pohon klasifikasi yang optimal.
7. Menghitung ketepatan klasifikasi pohon hasil bentukan dengan menggunakan data *learning* dan mengevaluasinya dengan menggunakan data *testing* pada pohon klasifikasi yang terbentuk. Ukuran ketepatan klasifikasi yang digunakan adalah *sensitivity*, *specificity* dan *total accuracy rate* (1-APER).

Langkah – langkah dari analisis yang digunakan digambarkan pada diagram alir (*flowchart*) sebagai berikut.





Gambar 3.1 *Flowchart* Langkah Analisis Penelitian

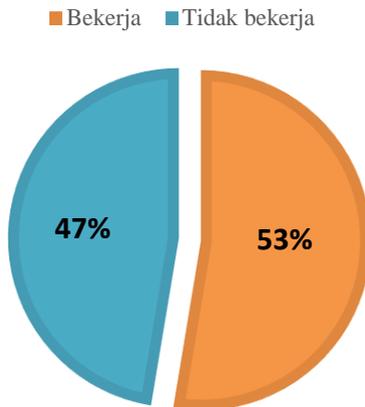
Halaman ini sengaja dikosongkan

BAB IV ANALISIS DAN PEMBAHASAN

Pada bab ini membahas mengenai hasil analisis data yang telah dilakukan untuk menjawab permasalahan dari penelitian. Analisis yang dilakukan yaitu analisis *Classification And Regression Trees* (CART) untuk mendapatkan hasil pengklasifikasian kegiatan ekonomi lansia yang dilihat dari bekerja atau tidak bekerja di Provinsi Jawa Timur. Sebelum dilakukan analisis CART terlebih dahulu mengetahui deskripsi karakteristik partisipasi kegiatan ekonomi berdasarkan faktor yang mempengaruhinya.

4.1 Deskripsi Karakteristik Partisipasi Kegiatan Ekonomi Lansia di Provinsi Jawa Timur

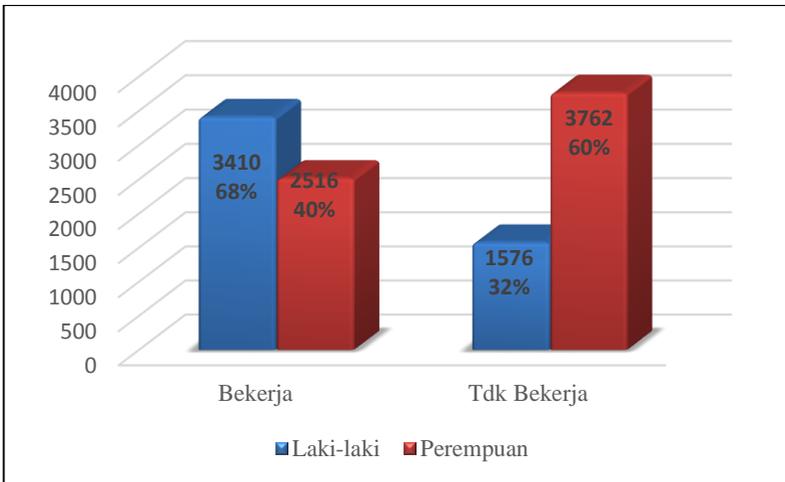
Untuk mendapatkan gambaran yang menyeluruh mengenai profil kegiatan ekonomi lansia yang terdiri dari lansia bekerja atau lansia tidak bekerja di provinsi Jawa Timur, maka perlu dilakukan deskripsi secara statistik dan disajikan pada diagram di bawah ini,



Gambar 4.1 Persentase Antara Lansia Bekerja dan Lansia Tidak Bekerja

Gambar 4.1 menunjukkan bahwa 53 persen atau sebanyak 5926 orang lansia dari jumlah keseluruhan lansia di provinsi Jawa Timur yaitu sebanyak 11264 lansia merupakan lansia yang bekerja atau ikut dalam partisipasi kegiatan ekonomi. Lansia yang tidak ikut dalam partisipasi ekonomi atau tidak bekerja memiliki persentase sebesar 47 persen atau sebanyak 5338 orang. Dari hasil tersebut dapat dilihat bahwa lansia yang ikut dalam partisipasi kegiatan ekonomi lebih tinggi daripada yang tidak berpartisipasi dalam kegiatan ekonomi.

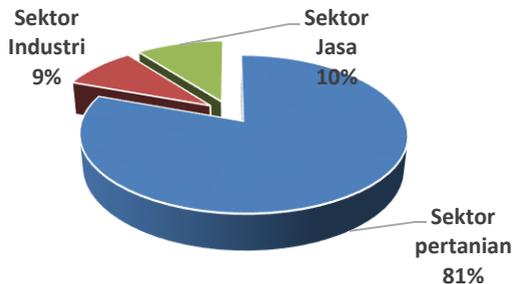
Gambar 4.2 dibawah menunjukkan hasil deskripsi lansia dengan status bekerja dan tidak bekerja berdasarkan jenis kelamin lansia.



Gambar 4.2 Persentase Lansia Bekerja dan Tidak Bekerja Berdasarkan Jenis Kelamin

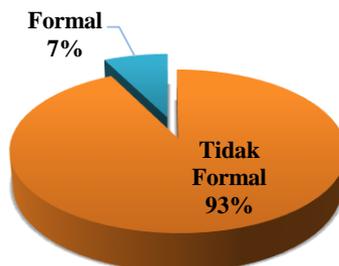
Gambar 4.2 menggambarkan bahwa lansia yang berjenis kelamin laki-laki dengan status bekerja memiliki persentase sebesar 68 persen atau sebanyak 3410 orang sedangkan untuk lansia yang berjenis kelamin perempuan sebesar 40 persen atau sebanyak 2516 orang. Dengan demikian lansia yang berjenis kelamin laki-laki dengan lebih tinggi persentasenya daripada lansia yang memiliki jenis kelamin perempuan dengan status bekerja.

Lansia yang bekerja di Provinsi Jawa Timur memiliki lapangan usaha atau bidang pekerjaan yang berbeda-beda. Bidang pekerjaan yang terdiri dari sektor pertanian, sektor industri dan sektor jasa. Gambar 4.3 dibawah menunjukkan bidang pekerjaan lansia di Provinsi Jawa Timur.



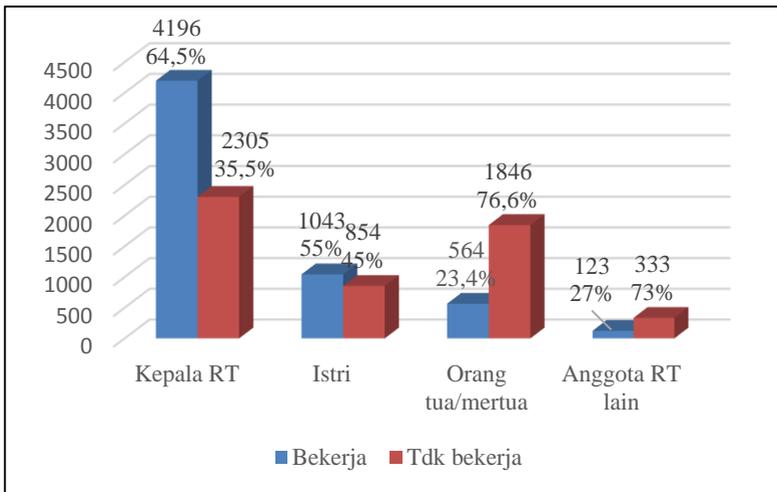
Gambar 4.3 Persentase Bidang Pekerjaan Lansia di Provinsi Jawa Timur

Berdasarkan Gambar 4.3 diatas diperoleh informasi bahwa lansia yang bekerja pada sektor pertanian memiliki persentase sebesar 81 persen atau sebanyak 4773 orang dari jumlah lansia yang berstatus kerja yaitu sebesar 5926 orang. Lansia yang bekerja pada sektor jasa sebesar 10 persen atau sebanyak 619 orang dan yang bekerja pada sektor industri sebesar 9 persen. Dari hasil diatas diperoleh bahwa para lansia di Provinsi Jawa Timur banyak yang bekerja pada sektor pertanian.



Gambar 4.4 Persentase Status/Kedudukan Lansia dalam Pekerjaan Utama

Gambar 4.4 diatas diperoleh informasi bahwa lansia yang bekerja memiliki dua status dalam pekerjaan utamanya, yaitu pekerjaan tidak formal dan pekerjaan formal. Persentase penduduk lansia yang memiliki status tidak formal dalam pekerjaannya sebesar 93 persen sedangkan sisanya sebesar 7 persen lansia yang memiliki status formal dalam pekerjaannya. Hal ini menandakan bahwa dari segi pendidikan lansia di Provinsi Jawa Timur tergolong rendah, karena itu lansia di Provinsi Jawa Timur sebagian besar memiliki status atau kedudukan tidak formal dalam pekerjaan utamanya.

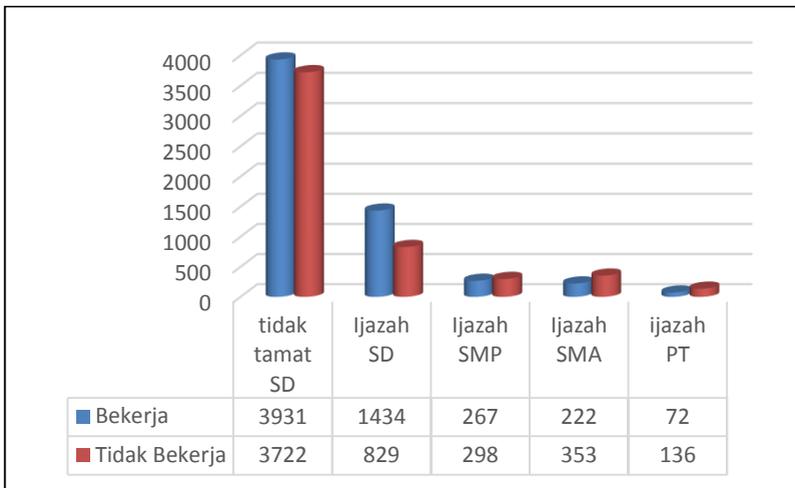


Gambar 4.5 Karakteristik Status Rumah tangga Lansia yang Bekerja dan Tidak Bekerja

Gambar 4.5 diatas menunjukkan jumlah lansia berdasarkan statusnya di rumah tangga yang bekerja dan tidak bekerja. Dari hasil pengolahan yang dilakukan diperoleh informasi bahwa sebanyak 4196 lansia atau sebesar 64,5 persen lansia yang bekerja di Provinsi Jawa Timur dengan berstatus sebagai kepala rumah tangga sedangkan lansia yang tidak bekerja dengan berstatus sebagai kepala rumah tangga sebanyak 2305 orang atau sebesar 35,5 persen. Lansia yang memiliki status rumah tangga sebagai istri dan bekerja sebanyak 1043 orang atau sebesar 55 persen sedangkan yang tidak bekerja sebanyak

854 orang atau sebesar 45 persen. Lansia yang bekerja sebanyak 564 orang atau sebesar 23,4 persen dengan status dalam rumah tangga sebagai orang tua/mertua. Sedangkan sisanya tidak bekerja. Sebanyak 123 orang lansia atau sebesar 27 persen di Provinsi Jawa Timur yang bekerja dengan status di dalam rumah tangga sebagai anggota rumah tangga lain sedangkan sisanya sebesar 73 persen lansia tidak bekerja dengan status sebagai anggota rumah tangga lain.

Gambar 4.6 dibawah merupakan deskripsi dari ijazah tertinggi yang dimiliki oleh lansia berdasarkan status bekerja dan status tidak bekerja.

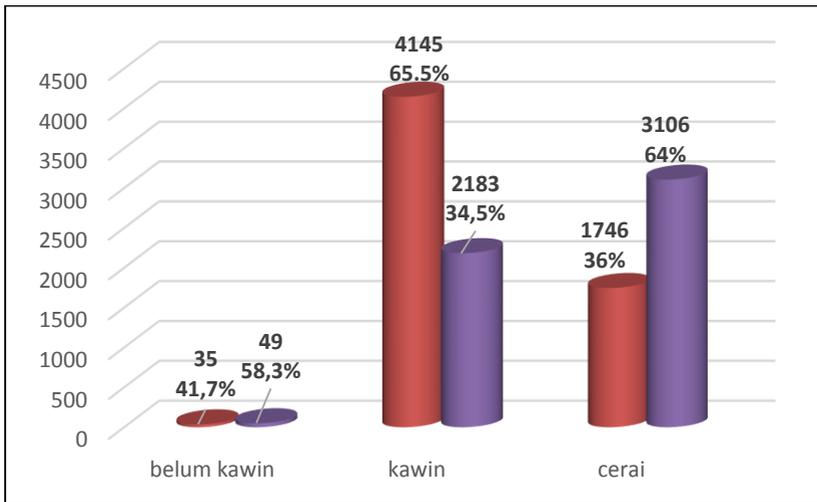


Gambar 4.6 Karakteristik Ijazah Tertinggi Lansia yang Bekerja dan Tidak Bekerja

Berdasarkan Gambar 4.6 diperoleh informasi bahwa lansia yang bekerja dengan ijazah tidak tamat SD sebanyak 3931 lansia, memiliki ijazah SD sebanyak 1434, memiliki ijazah SMP sebanyak 267, memiliki ijazah SMA sebanyak 222 dan yang memiliki ijazah perguruan tinggi sebanyak 7 lansia. Dari hasil diatas lansia yang berstatus bekerja dengan jumlah paling tinggi yaitu lansia yang tidak tamat SD, hal ini didukung oleh hasil analisis pada Gambar 4.4 yaitu

lansia yang bekerja sebagian besar memiliki status atau kedudukan tidak formal dalam pekerjaan utamanya. Gambar 4.6 diatas juga menunjukkan bahwa sebanyak 3722 lansia di Provinsi Jawa Timur tergolong lansia yang tidak bekerja dengan berstatus tidak tamat SD. Informasi selanjutnya dapat dilihat pada Gambar 4.6

Gambar 4.7 dibawah merupakan deskripsi dari status perkawinan lansia berdasarkan status bekerja dan tidak bekerja.

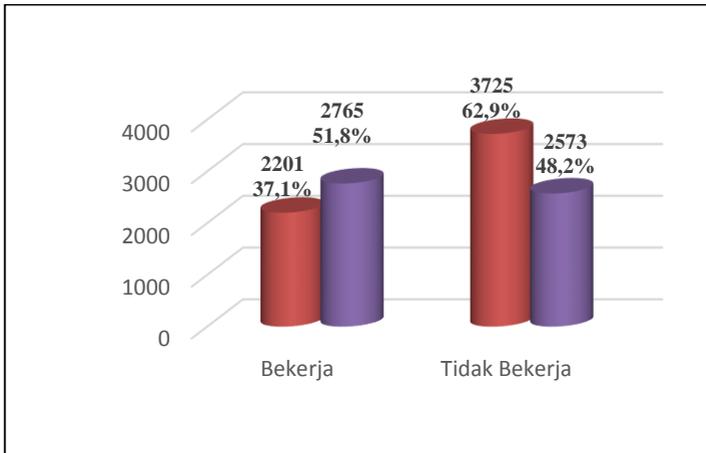


Gambar 4.7 Karakteristik Status Perkawinan Lansia yang Bekerja dan Tidak Bekerja

Gambar 4.7 diatas menunjukkan lansia yang belum kawin dengan status bekerja memiliki persentase sebesar 41,7 persen sedangkan sisanya sebesar 58,3 persen dengan status tidak bekerja. Lansia yang memiliki status perkawinan kawin dan tergolong lansia yang bekerja memiliki persentase sebesar 65,5 persen dan sisanya sebesar 34,5 persen tergolong lansia yang tidak bekerja. Terdapat sebesar 36 persen lansia yang status perkawinannya cerai hidup/cerai mati dengan kategori lansia yang bekerja sedangkan sisanya sebesar 64 persen tergolong lansia yang tidak bekerja. Berdasarkan jenis status perkawinannya dari jumlah total lansia sebanyak 11264,

sebanyak 52,6 persen atau sebanyak 5926 lansia yang tergolong berstatus bekerja, sisanya sebesar 47,4 persen atau sebanyak 5338 tergolong lansia yang tidak bekerja.

Gambar dibawah merupakan deskripsi dari klasifikasi tempat tinggal lansia (perkotaan/pedesaan) berdasarkan status bekerja dan tidak bekerja.

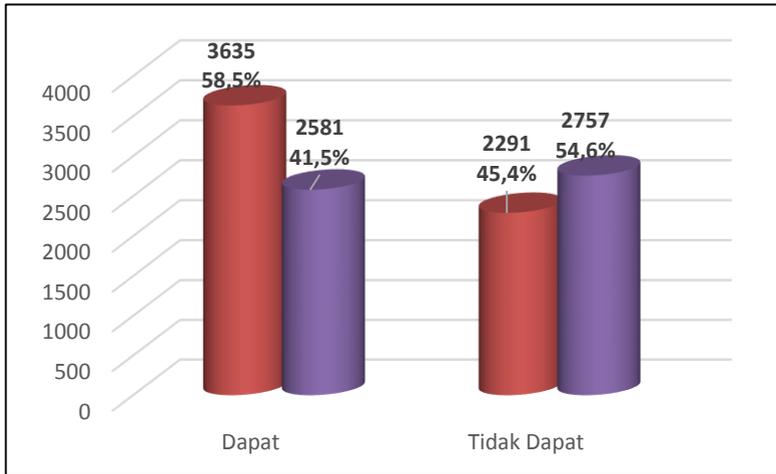


Gambar 4.8 Karakteristik Klasifikasi Perkotaan dan Pedesaan dengan Status Lansia Bekerja dan Tidak Bekerja

Gambar 4.8 diatas merupakan gambaran atau deskripsi yang memberikan informasi mengenai persentase lansia yang bekerja di daerah perkotaan dan di pedesaan. Di daerah perkotaan terdapat sebesar 37 persen atau sebanyak 2201 lansia, sedangkan di daerah pedesaan persentase lansia yang bekerja sebesar 63 persen atau sebanyak 3725 orang. Dapat dilihat dari kedua daerah tersebut yang memiliki persentase lebih tinggi yaitu terdapat pada daerah pedesaan

Gambar 4.9 dibawah memberikan informasi bahwa lansia yang dapat membaca dan menulis huruf latin tergolong lansia yang bekerja memiliki persentase sebesar 58,5 persen dan sisanya sebesar 41,5 persen lansia yang dapat membaca dan menulis tergolong lansia yang tidak bekerja. Untuk lansia yang tidak dapat membaca dan

menulis huruf latin tergolong lansia yang bekerja memiliki persentase sebesar 45,4 persen dan sisanya sebesar 54,6 persen lansia tergolong tidak bekerja. Berdasarkan kategori lansia yang dapat membaca dan menulis huruf latin atau tidak, dari total keseluruhan lansia yaitu sebanyak 11264 terdapat 52,6 persen tergolong lansia yang bekerja dan sebesar 47,4 persen tergolong lansia yang tidak bekerja.



Gambar 4.9 Karakteristik Lansia dapat Membaca dan Menulis Huruf Latin atau Tidak dengan Status Lansia Bekerja dan Tidak Bekerja

Tabel 4.1 Karakteristik Usia Lansia & Lamanya kesehatan Terganggu

Variabel	Rata-rata	Maksimum	Minimum	Variansi
Umur	69,29	98	60	57,75
Lamanya Kesehatan Terganggu	2,43	30	0	43,08

Tabel 4.1 diatas memberikan informasi mengenai karakteristik usia lansia dan lamanya lansia mengalami keluhan kesehatan di Provinsi Jawa Timur. Pada variabel umur lansia, dari 11264 lansia memiliki nilai rata-rata sebesar 69,29. Umur maksimum lansia yaitu lansia dengan umur 98 tahun dan umur lansia yang paling minimum yaitu umur 60 tahun. Variabel lamanya keluhan kesehatan lansia, dari total keseluruhan lansia sebanyak 11264 yaitu memiliki nilai rata-rata

sebesar 2,43 dengan lansia yang mengeluh mengenai kesehatannya yang paling tinggi lansia terganggu kesehatannya yaitu selama 30 hari dan nilai minimum dari variabel ini adalah 0 yang artinya lansia tidak mengalami keluhan kesehatan sama sekali selama 1 bulan terakhir.

4.2 Analisis CART untuk Klasifikasi Partisipasi Kegiatan Ekonomi Lansia di Provinsi Jawa Timur

Dalam analisis CART terdapat beberapa langkah, yaitu penentuan data *learning* dan data *testing*, penentuan pohon klasifikasi maksimal, pemangkasan pohon klasifikasi maksimal, penentuan pohon klasifikasi optimal dan validasi pohon klasifikasi. Berikut dari hasil analisis CART.

4.2.1 Penentuan Kombinasi Data Learning dan Data Testing

Kombinasi antara data *learning* dan data *testing* yang digunakan dalam penelitian ini yaitu terdapat 5 kombinasi data. Kelima kombinasi data *learning* dan *testing* tersebut yaitu (95 persen : 5 persen), (90 persen : 10 persen), (85 persen : 15 persen), (80 persen : 20 persen) dan (75 persen : 25 persen). Hasil klasifikasi yang diperoleh dari masing-masing kombinasi kemudian dilakukan perbandingan ketepatan klasifikasinya. Kombinasi proporsi data *learning* dan *testing* terbaik dipilih berdasarkan nilai ketepatan klasifikasinya yang paling besar. Data *learning* digunakan untuk mengsetimasi parameter model dan data *testing* digunakan untuk mencerminkan kebaikan model yang berupa pohon klasifikasi dalam mengklasifikasikan data baru.

Tabel 4.2 dibawah ini menunjukkan bahwa nilai ketepatan klasifikasi dari data *learning* dan data *testing* yang tertinggi dicapai pada kombinasi data *learning* 85 persen dan data *testing* 15 persen. Sehingga kombinasi data *learning* dan *testing* inilah yang akan digunakan untuk analisis selanjutnya. Analisis selanjutnya dengan menggunakan data *learning* 85 persen dan data *testing* 15 persen dilakukan pada langkah-langkah analisis CART.

Tabel 4.2 Perbandingan Ketepatan Klasifikasi Antar Kombinasi data

No	Kombinasi Data (%)		Ketepatan Klasifikasi	
	Learning	Testing	Learning	Testing
1	75	25	0,741	0,733
2	80	20	0,755	0,732
3	85	15	0,743	0,754
4	90	10	0,751	0,747
5	95	5	0,794	0,712

4.2.2 Pembentukan Pohon Klasifikasi Maksimal

Langkah pertama dalam pembentukan pohon klasifikasi maksimal yaitu dengan menentukan variabel pemilah dan nilai variabel (*threshold*). Variabel pemilah dan *threshold* didapatkan dari kemungkinan pemilah dari masing-masing variabel. Perhitungan banyaknya pemilah dari setiap variabel terdapat pada Tabel 4.3.

Tabel 4.3 Perhitungan Kemungkinan Pemilah dari Setiap Variabel

Nama Variabel	Skala Data	Banyaknya Kategori	Kemungkinan Pemilah
Jenis Kelamin (X_1)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Umur (X_2)	Rasio	39	$39 - 1 = 38$ pemilah
Hubungan dengan Kepala RT (X_3)	Nominal	4	$2^{4-1} - 1 = 7$ pemilah
Ijazah Tertinggi yang Dimiliki (X_4)	Ordinal	5	$5 - 1 = 4$ pemilah
Status Perkawinan (X_5)	Nominal	3	$2^{3-1} - 1 = 3$ pemilah

Tabel 4.3 Lanjutan Perhitungan Kemungkinan Pemilah dari Setiap Variabel

Nama Variabel	Skala Data	Banyaknya Kategori	Kemungkinan Pemilah
Klasifikasi Desa/Kel (X_6)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Dapat Membaca & Menulis Huruf Latin (X_7)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Keluhan Kesehatan selama berapa hari (X_8)	Interval	29	$29 - 1 = 28$ pemilah

Tabel 4.3 menunjukkan kemungkinan-kemungkinan pemilah yang terbentuk. Dari berbagai kemungkinan pemilah dari setiap variabel, selanjutnya dengan kriteria pemilihan (*goodness of split*) yang dihitung dengan Indeks Gini karena Indeks Gini dinilai lebih efektif dan sederhana untuk diaplikasikan. Cara kerja Indeks Gini adalah melakukan pemilahan simpul dengan berfokus pada masing-masing simpul kanan atau kiri. Pemilah yang terpilih adalah variabel pemilah dan nilai variabel (*threshold*) dengan *goodness of split* yang tertinggi. Pemilah yang terpilih merupakan variabel yang terpenting dalam mengklasifikasikan data pengamatan. Besarnya kontribusi variabel sebagai pemilah pada pohon klasifikasi maksimal yang terbentuk ditunjukkan melalui suatu angka skor yang ditampilkan pada Tabel 4.4.

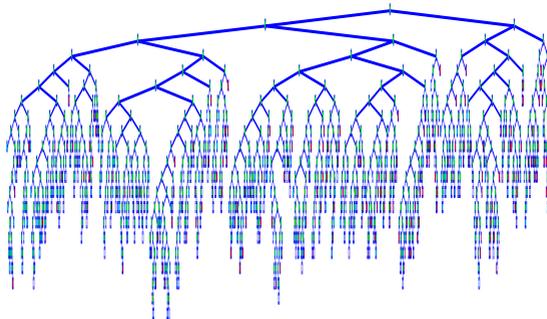
Berdasarkan Tabel 4.4 dibawah ini memberikan informasi bahwa semua variabel menjadi pembangun pohon klasifikasi. Variabel X_3 yaitu hubungan dengan kepala rumah tangga merupakan variabel yang terpenting atau variabel pemilah utama disebut juga dengan simpul induk (*parent node*) dalam mengklasifikasikan partisipasi kegiatan ekonomi lansia di Provinsi Jawa Timur, karena memiliki skor yang paling tinggi yaitu sebesar 100 persen. Selain itu variabel lain yang

memberikan pengaruh dalam melakukan pemilahan yaitu umur (X_2) memberikan kontribusi sebesar 66,93 persen ijazah tertinggi yang dimiliki (X_4) berkontribusi sebesar 34,16 persen, lamanya kesehatan terganggu dalam satu bulan terakhir (X_8) berkontribusi sebesar 33,47 persen dan variabel pembentuk lain selengkapnya dapat dilihat pada Tabel 4.4.

Tabel 4.4 Skor Variabel Terpenting pada Pohon Klasifikasi Maksimal

Variabel	Skor Variabel (%)	
X3	100,00	
X2	66,93	
X4	34,16	
X8	33,47	
X1	32,42	
X5	9,81	
X6	8,60	
X7	2,77	

Hasil penyekatan rekursif secara biner dari data pengamatan yang digunakan akan menghasilkan pohon klasifikasi yang relatif besar dan tingkat kedalaman (*depth*) yang tinggi. Pohon klasifikasi tersebut merupakan suatu pohon klasifikasi maksimal. Pada Gambar 4.10 dibawah disajikan pohon klasifikasi maksimal yang berukuran besar dengan tingkat kedalaman sebesar 44 dan jumlah *terminal nodes* sebanyak 730 *nodes*.

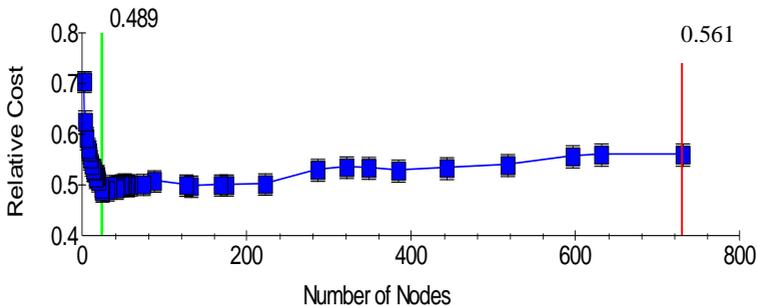


Gambar 4.10 Pohon Klasifikasi Maksimal Partisipasi Kegiatan Ekonomi Lansia Provinsi Jawa Timur

4.2.3 Pemangkasan Pohon Klasifikasi Maksimal (*Pruning*)

Untuk memudahkan proses analisis selanjutnya pohon klasifikasi maksimal yang sudah dihasilkan kemudian dilakukan pemangkasan pohon berdasarkan kriteria tertentu. Setiap hasil pemangkasan memiliki nilai *relative cost* tertentu. Sehingga kemudian dipilih nilai *relative cost* yang minimum.

Gambar 4.11 merupakan hasil pemangkasan dengan menggunakan metode *test sample estimate*. *Test sample estimate* lebih sesuai digunakan untuk pemangkasan dalam jumlah pengamatan yang besar. Algoritma dari *test sample estimate* yaitu membagi data penelitian menjadi dua bagian, yaitu L_1 untuk data *learning* dan L_2 untuk data *testing*. Selanjutnya L_1 digunakan untuk membentuk pohon T sedangkan data *testing* pada L_2 digunakan untuk menduga nilai *resubstitution estimate*. Pada Gambar 4.11 di bawah ini pohon klasifikasi maksimal ditunjukkan oleh garis berwarna merah dan pohon klasifikasi optimal ditunjukkan oleh garis berwarna hijau.



Gambar 4.11 Plot *Relative Cost* dan Jumlah *Terminal Nodes* dalam Klasifikasi Partisipasi Kegiatan Ekonomi Lansia

Urutan pembentukan pohon klasifikasi ditunjukkan pada Tabel 4.5 yang berisikan nilai simpul terminal, *test set relative cost*, *resubstitution relative cost* dan nilai *complexity parameter*.

Tabel 4.5 Urutan Pembentukan Pohon Klasifikasi

<i>Tree</i>	<i>Terminal Nodes</i>	<i>Test Set Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity Parameter</i>
1*	730	0,561 ± 0,022	0,407	0,000
27**	24	0,489 ± 0,021	0,513	0,417643E-03
33	14	0,529 ± 0,022	0,537	0,002
34	13	0,531 ± 0,021	0,542	0,002
35	11	0,541 ± 0,022	0,553	0,003
36	9	0,553 ± 0,022	0,565	0,003
37	8	0,566 ± 0,022	0,577	0,006
38	7	0,572 ± 0,022	0,590	0,007
39	5	0,594 ± 0,022	0,618	0,007
40	3	0,626 ± 0,022	0,653	0,009
41	2	0,705 ± 0,021	0,708	0,027
42	1	1,000 ± 0,000	1,000	0,146

Keterangan :

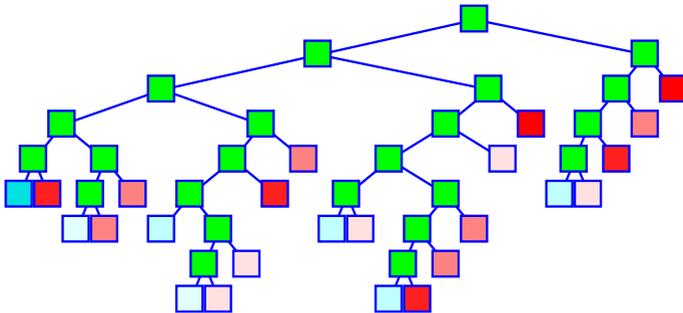
*) *Terminal nodes* dimana pohon maksimal terbentuk

***) *Terminal nodes* dimana pohon optimal terbentuk

Berdasarkan Tabel 4.5 dapat diketahui bahwa *terminal nodes* yang terbentuk pada pohon klasifikasi maksimal yaitu sebanyak 730 *terminal nodes* dan menghasilkan *complexity parameter* sebesar 0,000, *test set relative cost* sebesar $0,561 \pm 0,022$, dan *resubstitution relative cost* sebesar 0,407.

4.2.4 Pembentukan Pohon Klasifikasi Optimal

Hasil pemangkasan yang diperoleh dari Gambar 4.11 di atas selanjutnya digunakan untuk memilih pohon klasifikasi optimal. Pemilah yang terpilih sebagai simpul induk atau *parent node* dilihat dari hasil skor tertinggi pada setiap variabel prediktor. Berdasarkan informasi pada Tabel 4.5 pohon klasifikasi optimal terbentuk dengan jumlah *terminal nodes* sebanyak 24 simpul dengan nilai *complexity parameter* sebesar 0,417643E-03, *test set relative cost* yang paling minimum yaitu sebesar $0,489 \pm 0,021$ dan *resubstitution relative cost* sebesar 0,513.



Gambar 4.12 Pohon Klasifikasi Optimal untuk Partisipasi Kegiatan Ekonomi Lansia

Pohon klasifikasi optimal disajikan dalam Gambar 4.12 diatas. Pada Gambar 4.12 diatas pohon klasifikasi optimal dengan tingkat kedalaman pohon sebesar 9. Setiap *terminal nodes* yang terbentuk pada pohon klasifikasi optimal sudah dilakukan pelabelan kelas. Pelabelan kelas yang terbentuk terdiri dari dua warna, yaitu bewarna merah bila suatu simpul diklasifikasikan sebagai lansia yang tidak bekerja dan bewarna biru bila suatu simpul diklasifikasikan sebagai lansia yang bekerja. Apabila semakin kuat warna yang dihasilkan pada masing-masing simpul, hal tersebut mengindikasikan bahwa perbedaan proporsi antara kedua kategori partisipasi kegiatan ekonomi semakin besar dan nantinya simpul yang warnanya paling kuat yang akan dilakukan intepretasi karakteristik simpul.

Berdasarkan pohon klasifikasi optimal yang telah terbentuk, dengan informasi simpul yang lebih terinci pada Lampiran F, diperoleh bahwa variabel hubungan dengan kepala rumah tangga (X_3) merupakan variabel pemilah utama (*parent nodes*) dan paling menentukan klasifikasi partisipasi kegiatan ekonomi lansia dengan skor variabel sebesar 100,00. Hasil skor variabel pembentukan pohon klasifikasi optimal ditunjukkan pada Tabel 4.6 berikut.

Tabel 4.6 Skor Variabel Terpenting pada Pembentukan Pohon Klasifikasi Optimal

Variabel	Skor Variabel (%)	
X ₃	100,00	
X ₂	52,76	
X ₁	36,13	
X ₈	33,59	
X ₄	28,62	
X ₆	9,59	
X ₅	0,00	
X ₇	0,00	

Berdasarkan Tabel 4.6 diatas diketahui bahwa juga terdapat variabel lain selain variabel hubungan dengan kepala rumah tangga (X₃) yang berkontribusi terhadap pembentukan pohon klasifikasi optimal. Variabel umur (X₂) berkontribusi sebesar 52,76 persen pada pembentukan pohon klasifikasi optimal, variabel (X₁) jenis kelamin berkontribusi sebesar 36,13 persen, variabel lamanya lansia mengalami gangguan kesehatan (X₈) berkontribusi sebesar 33,59 persen, variabel ijazah tertinggi yang dimiliki lansia (X₄) berkontribusi sebesar 28,62 persen dan variabel klasifikasi desa/kelurahan tempat tinggal (X₆) dengan kemampuan memilah data paling kecil yaitu sebesar 9,59 persen. Skor yang diperoleh oleh setiap variabel tersebut menunjukkan besarnya pengaruh dari variabel tersebut untuk memilah data sehomogen mungkin.

Pengamatan simpul utama (simpul 1) dipilah menjadi dua simpul anak berdasarkan variabel hubungan dengan kepala rumah tangga (X₃) yaitu simpul kiri (simpul 2) terdiri dari 7153 pengamatan dengan hubungan kepala rumah tangga yang berstatus sebagai kepala rumah tangga dan sebagai istri dengan dilabeli sebagai lansia yang bekerja karena proporsi terbesar dalam simpul tersebut adalah proporsi lansia yang bekerja yaitu sebesar 0,622. Simpul kanan (simpul 3) terdiri dari 2421 pengamatan dengan hubungan kepala rumah tangga yang berstatus sebagai orang tua atau mertua dan anggota rumah tangga lainnya dengan label kelas sebagai lansia yang tidak bekerja karena proporsi terbesar dalam simpul tersebut adalah proporsi lansia yang

tidak bekerja yaitu sebesar 0,763 (76,3 persen). Selanjutnya, setiap simpul dilakukan pemilahan lagi karena masih memiliki tingkat keheterogenan yang tinggi. Pemilahan dilakukan hingga diperoleh *terminal nodes* sebanyak 24 simpul terminal. Proses pemilahan hingga didapatkan *terminal nodes* sebanyak 24 simpul disajikan secara terperinci pada Lampiran F.

Setiap simpul terminal merupakan titik akhir dari suatu pemilahan, simpul ini sudah tidak dapat dipilah kembali karena sudah homogen. Simpul terminal yang terbentuk pada penelitian ini adalah sebanyak 24 simpul terminal, sebanyak 7 simpul terminal yang diklasifikasikan sebagai lansia dengan kategori bekerja dan sebanyak 17 simpul terminal diklasifikasikan sebagai kategori lansia yang tidak bekerja. Dari hasil simpul terminal yang terbentuk, karakteristik dari beberapa simpul terminal yang paling kuat mengklasifikasikan data menjadi salah satu dari kategori variabel Y yaitu partisipasi kegiatan ekonomi lansia dengan dilihat nilai persentase yang terbesar dari hasil klasifikasi setiap simpul terminal ke dalam salah satu kategori variabel Y.

1. **Simpul terminal 1** terdiri dari 2021 anggota lansia yang dilabeli sebagai kelas 1 yaitu sebagai lansia yang bekerja. Karakteristik lansia pada simpul ini adalah lansia dengan status sebagai kepala rumah tangga dan sebagai istri, lansia yang mengalami keluhan kesehatan kurang dari samadengan 27 hari dengan pendidikan terakhir lansia tidak tamat SD, ijazah tertinggi SD dan ijazah tertinggi SMP/ sederajat, tergolong lansia yang berjenis kelamin laki-laki dengan umur lansia tidak melebihi 69,5 tahun.
2. **Simpul terminal 2** terdiri dari 70 anggota lansia yang dilabeli sebagai kelas 2 yaitu lansia yang tidak bekerja. Karakteristik lansia pada simpul terminal 2 ini adalah lansia dengan status sebagai kepala rumah tangga dan sebagai istri, lansia yang mengalami keluhan kesehatan lebih dari 27 hari dengan pendidikan terakhir lansia tidak tamat SD, ijazah tertinggi SD, ijazah tertinggi SMP/ sederajat dan tergolong lansia yang berjenis kelamin laki-laki dengan umur lansia tidak melebihi 69,5 tahun.

3. **Simpul terminal 10** terdiri dari 89 anggota lansia yang dilabeli sebagai kelas 2 yaitu lansia yang tidak bekerja. Karakteristik lansia pada simpul terminal 10 ini adalah lansia yang berstatus sebagai kepala rumah tangga dan sebagai istri, lansia yang mengalami keluhan kesehatan selama lebih dari 7,5 hari dengan ijazah tertinggi yang dimiliki lansia adalah tidak tamat SD dan ijazah tertinggi SD dengan tergolong lansia yang berjenis kelamin perempuan dengan umur lansia tidak melebihi 69,5 tahun.
4. **Simpul terminal 12** terdiri dari 515 anggota lansia yang dilabeli sebagai kelas 1 yaitu lansia yang mempunyai kategori bekerja. Karakteristik lansia pada simpul terminal 12 ini adalah lansia yang berstatus sebagai kepala rumah tangga dan sebagai istri, lansia yang berumur antara 69 tahun hingga 78,5 tahun dengan lansia yang bertempat tinggal di daerah pedesaan dan tergolong lansia yang memiliki jenis kelamin laki-laki serta lansia yang pernah mengalami gangguan kesehatan selama kurang dari 13,5 hari.
5. **Simpul terminal 15** terdiri dari 21 anggota lansia yang dilabeli sebagai kelas 2 yaitu lansia yang mempunyai kategori tidak bekerja. Karakteristik lansia pada simpul terminal 15 ini adalah lansia yang berstatus sebagai kepala rumah tangga dan sebagai istri, lansia yang mengalami keluhan kesehatan selama rentang waktu 6,5 hari sampai dengan 13,5 hari. Ijazah tertinggi yang dimiliki lansia tidak tamat SD dan ijazah tertinggi SD, umur lansia pada kelompok ini yaitu tidak lebih dari 77,5 tahun dan lansia yang bertempat tinggal di wilayah perkotaan dengan tergolong lansia yang berjenis kelamin laki-laki.
6. **Simpul terminal 19** terdiri dari 204 anggota lansia yang dilabeli sebagai kelas 2 yaitu lansia yang mempunyai kategori tidak bekerja. Karakteristik lansia pada simpul ini adalah lansia yang berstatus sebagai kepala rumah tangga dan istri, lansia yang mengalami keluhan kesehatan selama lebih dari 13,5 hari dengan lansia yang tergolong umur lebih dari 69,5 tahun.

7. **Simpul terminal 22** terdiri dari 41 anggota lansia yang dilabeli sebagai kelas 2 yaitu lansia yang memiliki kategori lansia tidak bekerja. Karakteristik lansia pada simpul ini adalah lansia yang memiliki status sebagai orang tua/mertua dan sebagai anggota rumah tangga lainnya, lansia yang mengalami keluhan kesehatan selama lebih dari 3,5 hari dengan lansia yang tergolong memiliki jenis kelamin laki-laki dan lansia berumur tidak lebih dari 72,5 tahun.
8. **Simpul terminal 23** terdiri dari 1022 anggota lansia yang dilabeli sebagai kelas 2 yaitu lansia yang tidak bekerja. Karakteristik lansia pada simpul terminal 23 ini adalah lansia yang berstatus sebagai orang tua/mertua dan sebagai anggota rumah tangga lainnya, lansia yang memiliki jenis kelamin perempuan dan tergolong lansia yang berumur tidak lebih dari 72,5 tahun.
9. **Simpul terminal 24** terdiri dari 1147 anggota lansia yang dilabeli sebagai kelas 2 yaitu lansia yang tidak bekerja. Karakteristik lansia pada simpul ini adalah lansia yang berstatus sebagai orang tua/mertua dan sebagai anggota rumah tangga lainnya dan tergolong lansia yang berumur lebih dari 72,5 tahun.

4.2.5 Hasil Ketepatan Klasifikasi CART

Tingkat keakuratan hasil klasifikasi pohon optimal dari data *learning* dan data *testing* dapat ditunjukkan pada Tabel 4.7 dan Tabel 4.8 berikut.

Tabel 4.7 Klasifikasi Pohon Optimal

Observasi		Prediksi		Total
		Bekerja	Tidak Bekerja	
Data	Bekerja	3633	1393	5026
Learning	Tidak Bekerja	1072	3476	4548
Total Tingkat Kesalahan (APER)				25,7%
Tingkat Akurasi Total (1-APER)				74,3%
<i>Sensitivity (%)</i>				72,3%
<i>Specificity (%)</i>				76,4%

Perhitungan akurasi, *sensitivity*, *specificity* untuk data *learning* adalah sebagai berikut.

$$\text{APER (\%)} = \left(\frac{1393+1072}{9574} \right) \times 100\% = 25,7\%$$

$$\text{Akurasi} = 1 - \text{APER} = 74,3\%$$

$$\text{Sensitivity} = \left(\frac{3633}{5026} \right) \times 100\% = 72,3\%$$

$$\text{Specificity} = \left(\frac{3476}{4548} \right) \times 100\% = 76,4\%$$

Berdasarkan perhitungan yang dilakukan diatas, dapat diambil kesimpulan bahwa dari hasil pembentukan pohon klasifikasi optimal Nilai tingkat akurasi total pada data *learning* sebesar 74,3 persen yang artinya pohon klasifikasi yang terbentuk mampu memprediksi dengan tepat pengamatan sebesar 74,3 persen. Tingkat akurasi pada lansia yang tergolong bekerja ditunjukkan oleh nilai *sensitivity* sebesar 72,3 persen dan tingkat akurasi pada lansia yang tergolong tidak bekerja ditunjukkan oleh nilai *specificity* yaitu sebesar 76,4 persen.

Tabel 4.8 Validasi Hasil Klasifikasi Pohon Optimal

Observasi		Prediksi		Total
		Bekerja	Tidak Bekerja	
Data	Bekerja	654	246	900
Testing	Tidak Bekerja	170	620	790
Total Tingkat Kesalahan (APER)				24,6%
Tingkat Akurasi Total (1-APER)				75,4%
<i>Sensitivity (%)</i>				72,7%
<i>Specificity (%)</i>				78,5%

Apabila hasil klasifikasi yang terbentuk divalidasi dengan menggunakan data *testing* diperoleh hasil seperti disajikan pada Tabel 4.8 diatas bahwa sebanyak 246 lansia yang salah diklasifikasikan sebagai lansia yang tidak bekerja dan sebanyak 170 lansia salah diklasifikasikan sebagai lansia yang bekerja.

Perhitungan akurasi, *sensitivity*, *specificity* untuk data *testing* adalah sebagai berikut.

$$\text{APER (\%)} = \left(\frac{246+170}{1690} \right) \times 100\% = 24,6\%$$

$$\text{Akurasi} = 1 - \text{APER} = 75,4\%$$

$$\text{Sensitivity} = \left(\frac{654}{900} \right) \times 100\% = 72,7\%$$

$$\text{Specificity} = \left(\frac{620}{790} \right) \times 100\% = 78,5\%$$

Nilai tingkat akurasi total pada data *testing* sebesar 75,4 persen artinya model pohon yang telah terbentuk memiliki keakuratan hasil prediksi sebesar 75,4 persen. Nilai *sensitivity* yang terbentuk sebesar 72,7 persen yang artinya tingkat akurasi pada lansia yang tergolong bekerja yaitu sebesar 72,7 persen. Tingkat akurasi pada lansia yang tergolong tidak bekerja ditunjukkan oleh nilai *specificity* yaitu sebesar 78,5 persen. Nilai tingkat akurasi total untuk data *testing* tergolong tinggi, sehingga dapat dikatakan bahwa pohon optimal yang terbentuk sudah baik dan sesuai atau layak bila digunakan untuk mengklasifikasikan data baru.

Halaman sengaja dikosongkan

LAMPIRAN A

Data Pengamatan Lansia Berdasarkan Status Bekerja dan Tidak Bekerja

No	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
1	2	2	92	3	1	3	2	2	0
2	1	2	64	3	1	3	2	2	0
3	1	1	76	1	1	2	2	2	0
4	1	2	60	2	1	2	2	1	0
5	1	2	60	2	2	2	2	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11263	2	1	82	3	1	2	2	2	0
11264	2	2	66	3	1	2	2	2	0

Keterangan :

Y : Partisipasi kegiatan ekonomi lansia (bekerja/tidak bekerja)

X₁ : Jenis kelamin

X₂ : Umur

X₃ : Hubungan dengan kepala rumah tangga

X₄ : Ijazah tertinggi yang dimiliki

X₅ : Status perkawinan

X₆ : Klasifikasi desa/kelurahan tempat tinggal

X₇ : Dapat membaca dan menulis huruf latin

X₈ : Lamanya lansia mengalami gangguan kesehatan

LAMPIRAN B**Output CART untuk Data Learning 90% dan Data Testing 10%**

```
=====
TEST SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	439.00	150.00	589.00
2	135.00	402.00	537.00
PRED. TOT.	574.00	552.00	1126.00
CORRECT	0.745	0.749	
SUCCESS IND.	0.222	0.272	
TOT. CORRECT	0.747		

```
SENSITIVITY:      0.745      SPECIFICITY:      0.749
FALSE REFERENCE:  0.235    FALSE RESPONSE:   0.272
REFERENCE = "1", RESPONSE = "2"
```

```
=====
TEST SAMPLE CLASSIFICATION PROBABILITY TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.745	0.255	1.000
2	0.251	0.749	1.000

```
=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	4075.00	1262.00	5337.00
2	1261.00	3540.00	4801.00
PRED. TOT.	5336.00	4802.00	10138.00
CORRECT	0.764	0.737	

SUCCESS IND. 0.237 0.264
 TOT. CORRECT 0.751

SENSITIVITY: 0.764 SPECIFICITY: 0.737
 FALSE REFERENCE: 0.236 FALSE RESPONSE: 0.263
 REFERENCE = "1", RESPONSE = "2"

 =====
 LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE
 =====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.764	0.236	1.000
2	0.263	0.737	1.000

Output CART untuk Data Learning 95% dan Data Testing 5%

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	203.00	91.00	294.00
2	71.00	198.00	269.00
PRED. TOT.	274.00	289.00	563.00
CORRECT	0.690	0.736	
SUCCESS IND.	0.168	0.258	
TOT. CORRECT	0.712		

SENSITIVITY: 0.690 SPECIFICITY: 0.736
 FALSE REFERENCE: 0.259 FALSE RESPONSE: 0.315
 REFERENCE = "1", RESPONSE = "2"

 =====
 TEST SAMPLE CLASSIFICATION PROBABILITY TABLE
 =====

Actual Class	Predicted Class		Actual Total
	1	2	

1	0.690	0.310	1.000
2	0.264	0.736	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	4555.00	1077.00	5632.00
2	1123.00	3946.00	5069.00
PRED. TOT.	5678.00	5023.00	10701.00
CORRECT	0.809	0.778	
SUCCESS IND.	0.282	0.305	
TOT. CORRECT	0.794		

SENSITIVITY: 0.809 SPECIFICITY: 0.778
 FALSE REFERENCE: 0.198 FALSE RESPONSE: 0.214
 REFERENCE = "1", RESPONSE = "2"

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.809	0.191	1.000
2	0.222	0.778	1.000

Output CART untuk Data *Learning* 85% dan Data *Testing* 15%

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	654.00	246.00	900.00
2	170.00	620.00	790.00
PRED. TOT.	824.00	866.00	1690.00

CORRECT	0.727	0.785
SUCCESS IND.	0.194	0.317
TOT. CORRECT	0.754	

SENSITIVITY:	0.727	SPECIFICITY:	0.785
FALSE REFERENCE:	0.206	FALSE RESPONSE:	0.284
REFERENCE = "1", RESPONSE = "2"			

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.727	0.273	1.000
2	0.215	0.785	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	3633.00	1393.00	5026.00
2	1072.00	3476.00	4548.00
PRED. TOT.	4705.00	4869.00	9574.00
CORRECT	0.723	0.764	
SUCCESS IND.	0.198	0.289	
TOT. CORRECT	0.743		

SENSITIVITY:	0.723	SPECIFICITY:	0.764
FALSE REFERENCE:	0.228	FALSE RESPONSE:	0.286
REFERENCE = "1", RESPONSE = "2"			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.723	0.277	1.000
2	0.236	0.764	1.000

Output CART untuk Data Learning 80% dan Data Testing 20%

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	859.00	301.00	1160.00
2	302.00	791.00	1093.00
PRED. TOT.	1161.00	1092.00	2253.00
CORRECT	0.741	0.724	
SUCCESS IND.	0.226	0.239	
TOT. CORRECT	0.732		

SENSITIVITY: 0.741 SPECIFICITY: 0.724
 FALSE REFERENCE: 0.260 FALSE RESPONSE: 0.276
 REFERENCE = "1", RESPONSE = "2"

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.741	0.259	1.000
2	0.276	0.724	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	3638.00	1128.00	4766.00
2	1081.00	3164.00	4245.00
PRED. TOT.	4719.00	4292.00	9011.00
CORRECT	0.763	0.745	
SUCCESS IND.	0.234	0.274	
TOT. CORRECT	0.755		

SENSITIVITY: 0.763 SPECIFICITY: 0.745

FALSE REFERENCE: 0.229 FALSE RESPONSE: 0.263
 REFERENCE = "1", RESPONSE = "2"

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.763	0.237	1.000
2	0.255	0.745	1.000

Output CART untuk Data Learning 75% dan Data Testing 25%

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	1051.00	444.00	1495.00
2	307.00	1014.00	1321.00
PRED. TOT.	1358.00	1458.00	2816.00
CORRECT	0.703	0.768	
SUCCESS IND.	0.172	0.298	
TOT. CORRECT	0.733		

SENSITIVITY: 0.703 SPECIFICITY: 0.768
 FALSE REFERENCE: 0.226 FALSE RESPONSE: 0.305
 REFERENCE = "1", RESPONSE = "2"

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.703	0.297	1.000
2	0.232	0.768	1.000

```
=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	3182.00	1249.00	4431.00
2	940.00	3077.00	4017.00
PRED. TOT.	4122.00	4326.00	8448.00
CORRECT	0.718	0.766	
SUCCESS IND.	0.194	0.290	
TOT. CORRECT	0.741		

SENSITIVITY: 0.718 SPECIFICITY:
 0.766
 FALSE REFERENCE: 0.228 FALSE RESPONSE:
 0.289
 REFERENCE = "1", RESPONSE = "2"

```
-----
=====
LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.718	0.282	1.000
2	0.234	0.766	1.000

LAMPIRAN C

Output CART untuk Hasil Klasifikasi Menggunakan Proporsi Data yang Terpilih

Variable Importance

Variable	Score	
X3	100.00	
X2	66.93	
X4	34.16	
X8	33.47	
X1	32.42	
X5	9.81	
X6	8.60	
X7	2.77	

Misclassification for Learn Data

Class	N Cases	N Mis-classed	Pct Error	Cost
2	4548	1072	23.57	0.24
1	5026	1393	27.72	0.28

Misclassification for Test Data

Class	N Cases	N Mis-classed	Pct Error	Cost
2	790	170	21.52	0.22
1	900	246	27.33	0.27

Output Statistik Variabel Penelitian

LEARNING AND TEST (*) SAMPLE VARIABLE STATISTICS

```
=====
```

VARIABLE		CLASS		
		1	2	OVERALL
Y	MEAN	1.000	2.000	1.475
	SD	0.000	0.000	0.499
	N	5026.000	4548.000	9574.000
	SUM	5026.000	9096.000	14122.000
*	MEAN	1.000	2.000	1.467
*	SD	0.000	0.000	0.499
*	N	900.000	790.000	1690.000
*	SUM	900.000	1580.000	2480.000

X1	MEAN	1.423	1.703	1.556
	SD	0.494	0.457	0.497
	N	5026.000	4548.000	9574.000
	SUM	7152.000	7746.000	14898.000
	* MEAN	1.433	1.714	1.564
	* SD	0.496	0.452	0.496
	* N	900.000	790.000	1690.000
X2	* SUM	1290.000	1354.000	2644.000
	MEAN	66.864	71.978	69.293
	SD	6.086	8.180	7.599
	N	5026.000	4548.000	9574.000
	SUM	336058.000	327356.000	663414.000
	* MEAN	66.579	72.405	69.302
	* SD	5.783	8.223	7.606
X3	* N	900.000	790.000	1690.000
	* SUM	59921.000	57200.000	117121.000
	MEAN	1.422	2.036	1.713
	SD	0.743	1.012	0.933
	N	5026.000	4548.000	9574.000
	SUM	7145.000	9259.000	16404.000
	* MEAN	1.468	2.056	1.743
X4	* SD	0.777	1.018	0.944
	* N	900.000	790.000	1690.000
	* SUM	1321.000	1624.000	2945.000
	MEAN	1.496	1.561	1.527
	SD	0.846	1.022	0.934
	N	5026.000	4548.000	9574.000
	SUM	7521.000	7099.000	14620.000
X5	* MEAN	1.474	1.604	1.535
	* SD	0.825	1.073	0.951
	* N	900.000	790.000	1690.000
	* SUM	1327.000	1267.000	2594.000
	MEAN	2.289	2.570	2.422
	SD	0.467	0.513	0.509
	N	5026.000	4548.000	9574.000
X6	SUM	11503.000	11688.000	23191.000
	* MEAN	2.289	2.589	2.429
	* SD	0.461	0.513	0.508
	* N	900.000	790.000	1690.000
	* SUM	2060.000	2045.000	4105.000
	MEAN	1.626	1.486	1.560
	SD	0.484	0.500	0.496
X6	N	5026.000	4548.000	9574.000
	SUM	8172.000	6759.000	14931.000

	* MEAN	1.643	1.458	1.557
	* SD	0.479	0.499	0.497
	* N	900.000	790.000	1690.000
	* SUM	1479.000	1152.000	2631.000
X7	MEAN	1.385	1.520	1.449
	SD	0.487	0.500	0.497
	N	5026.000	4548.000	9574.000
	SUM	6962.000	6912.000	13874.000
	* MEAN	1.394	1.497	1.443
	* SD	0.489	0.500	0.497
	* N	900.000	790.000	1690.000
	* SUM	1255.000	1183.000	2438.000
X8	MEAN	1.118	3.872	2.427
	SD	3.542	8.500	6.542
	N	5026.000	4548.000	9574.000
	SUM	5620.000	17612.000	23232.000
	* MEAN	1.178	3.911	2.456
	* SD	3.667	8.741	6.686
	* N	900.000	790.000	1690.000
	* SUM	1060.000	3090.000	4150.000

Output Relative Cost

=====

TREE SEQUENCE

=====

Dependent variable: Y

Terminal	Test Set	Resubstitution	Complexity	
Tree Nodes	Relative Cost	Relative Cost	Parameter	
1	730	0.561 +/- 0.022	0.407	0.000
27**	24	0.489 +/- 0.021	0.513	.417643E-03
33	14	0.529 +/- 0.022	0.537	0.002
34	13	0.531 +/- 0.021	0.542	0.002
35	11	0.541 +/- 0.022	0.553	0.003
36	9	0.553 +/- 0.022	0.565	0.003
37	8	0.566 +/- 0.022	0.577	0.006
38	7	0.572 +/- 0.022	0.590	0.007
39	5	0.594 +/- 0.022	0.618	0.007
40	3	0.626 +/- 0.022	0.653	0.009
41	2	0.705 +/- 0.021	0.708	0.027
42	1	1.000 +/- 0.000	1.000	0.146

Output Informasi tiap-tiap Simpul Pohon Klasifikasi Optimal

```
*****
*           Node 1: X3           *
*           N: 9574             *
*****
```

```
*****
* Node 2 *           * Node 20 *
* N: 7153 *          * N: 2421 *
*****
```

Node 1 was split on X3

A case goes left if $X3 = (1,2)$

Improvement = 0.055 Complexity Threshold = 0.146

Node	Cases	Wgt Counts	Cost	Class
1	9574	9574.00	0.500	1
2	7153	7153.00	0.401	1
20	2421	2421.00	0.219	2

Class	Weighted Counts		
	Top	Left	Right
1	5026.00	4452.00	574.00
2	4548.00	2701.00	1847.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.500	0.599	0.219
2	0.500	0.401	0.781

Surrogate	Split	Assoc.	Improve.
1 X2	s 84.500	0.037	0.012

Competitor	Split	Improve.
1 X2	69.500	0.044
2 X5	2	0.042
3 X1	1	0.040
4 X8	14.500	0.018
5 X6	2	0.010

```
*****
*           Node 2: X2           *
*           N: 7153             *
*****
```

```
*****
* Node 3 *           * Node 13 *
* N: 4606 *         * N: 2547 *
*****
```

Node 2 was split on X2

A case goes left if X2 <= 69.500

Improvement = 0.019 Complexity Threshold = 0.027

Node	Cases	Wgt Counts	Cost	Class
2	7153	7153.00	0.401	1
3	4606	4606.00	0.316	1
13	2547	2547.00	0.449	2

Class	Weighted Counts		
	Top	Left	Right
1	4452.00	3246.00	1206.00
2	2701.00	1360.00	1341.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.599	0.684	0.449
2	0.401	0.316	0.551

Competitor	Split	Improve.
1 X1	1	0.018
2 X8	14.500	0.016
3 X6	2	0.013
4 X5	1,2	0.010
5 X4	1,2	0.008

```
*****
*           Node 3: X1           *
*           N: 4606             *
*****
```

```
*****
* Node 4 *           * Node 8 *
* N: 2417 *         * N: 2189 *
*****
```

Node 3 was split on X1
 A case goes left if X1 = (1)
 Improvement = 0.014 Complexity Threshold = 0.007

Node	Cases	Wgt Counts	Cost	Class
3	4606	4606.00	0.316	1
4	2417	2417.00	0.200	1
8	2189	2189.00	0.442	1

Weighted Counts

Class	Top	Left	Right
1	3246.00	1970.00	1276.00
2	1360.00	447.00	913.00

Within Node Probabilities

Class	Top	Left	Right
1	0.684	0.800	0.558
2	0.316	0.200	0.442

Surrogate	Split	Assoc.	Improve.
1 X3	s 1	0.573	0.006
2 X5	s 1,2	0.319	0.003
3 X7	s 1		
4 X4	s 2,3,4,5	0.185	0.003

Competitor	Split	Improve.
1 X4	1,2	0.010
2 X6	2	0.009
3 X8	24.500	0.009
4 X3	1	0.006
5 X5	1,2	0.003

```
*****
*                               *
*           Node 4: X4          *
*           N: 2417            *
*                               *
*****
```

```
*****
* Node 5 *                       * Node 6 *
* N: 2091 *                       * N: 326 *
*****
```

Node 4 was split on X4
 A case goes left if X4 = (1,2,3)
 Improvement = 0.008 Complexity Threshold = 0.003

Node	Cases	Wgt Counts	Cost	Class
4	2417	2417.00	0.200	1
5	2091	2091.00	0.149	1
6	326	326.00	0.481	2

Weighted Counts

Class	Top	Left	Right
1	1970.00	1805.00	165.00
2	447.00	286.00	161.00

Within Node Probabilities

Class	Top	Left	Right
1	0.800	0.851	0.481
2	0.200	0.149	0.519

Competitor	Split	Improve.
1 X8	27.000	0.007
2 X6	2	0.004
3 X2	61.500	.903025E-03
4 X7	2	.755283E-03
5 X5	1,2	.625413E-03


```

*****
*           Node 5: X8           *
*           N: 2091             *
*****
*****
* Node 1 *           * Node 2 *
* N: 2021 *           * N: 70  *
*****
Node 5 was split on X8
A case goes left if X8 <= 27.000
Improvement = 0.007      Complexity Threshold = 0.005
Node   Cases   Wgt Counts   Cost   Class
  5     2091     2091.00     0.149  1
 -1     2021     2021.00     0.124  1
 -2       70       70.00     0.184  2

```

Weighted Counts

Class	Top	Left	Right
1	1805.00	1791.00	14.00
2	286.00	230.00	56.00

Within Node Probabilities

Class	Top	Left	Right
1	0.851	0.876	0.184
2	0.149	0.124	0.816

Competitor	Split	Improve.
1 X6	2	0.001
2 X4	1,2	0.001
3 X5	1,2	.737418E-03
4 X2	61.500	.581302E-03
5 X7	2	.106366E-03

```

*****
*           Node 6: X2           *
*           N: 326              *
*****
*****
* Node 7 *           = Terminal Node 5 =
* N: 254 *           = N : 72         =
*****
*****

```

Node 6 was split on X2
 A case goes left if X2 <= 65.500
 Improvement = .989959E-03 Complexity Threshold = 0.002

Node	Cases	Wgt	Counts	Cost	Class
6	326		326.00	0.481	2
7	254		254.00	0.454	1
-5	72		72.00	0.258	2

Class	Weighted Counts		
	Top	Left	Right
1	165.00	145.00	20.00
2	161.00	109.00	52.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.481	0.546	0.258
2	0.519	0.454	0.742

Competitor	Split	Improve.
1 X6	2	.463834E-03
2 X8	14.500	.418015E-03
3 X4	4	.400706E-04
4 X5	1,2	.279887E-04

```
*****
*           Node 7: X8           *
*           N: 254               *
*****
```

```
=====
= Terminal Node 3 =           = Terminal Node 4 =
= N : 226         =           = N : 28         =
=====
```

Node 7 was split on X8

A case goes left if $X8 \leq 3.500$

Improvement = .531383E-03 Complexity Threshold = 0.001

Node	Cases	Wgt	Counts	Cost	Class
7	254		254.00	0.454	1
-3	226		226.00	0.418	1
-4	28		28.00	0.266	2

Weighted Counts

Class	Top	Left	Right
1	145.00	137.00	8.00
2	109.00	89.00	20.00

Within Node Probabilities

Class	Top	Left	Right
1	0.546	0.582	0.266
2	0.454	0.418	0.734

Competitor	Split	Improve.
1 X6	2	.231543E-03
2 X4	4	.111164E-03
3 X2	61.500	.110041E-03
4 X5	1,2	.638526E-04

```
*****
*           Node 8: X4           *
*           N: 2189              *
*****
```

```
*****
* Node 9 *           = Terminal Node 11 =
* N: 1915 *         = N : 274         =
*****
```

Node 8 was split on X4

A case goes left if $X4 = (1,2)$

Improvement = 0.006 Complexity Threshold = 0.014

Node	Cases	Wgt	Counts	Cost	Class
8	2189		2189.00	0.442	1
9	1915		1915.00	0.398	1
-11	274		274.00	0.265	2

Weighted Counts			
Class	Top	Left	Right
1	1276.00	1198.00	78.00
2	913.00	717.00	196.00
Within Node Probabilities			
Class	Top	Left	Right
1	0.558	0.602	0.265
2	0.442	0.398	0.735

Competitor	Split	Improve.
1 X6	2	0.005
2 X8	7.500	0.003
3 X7	2	0.001
4 X2	65.500	.826016E-03
5 X5	2,3	.365124E-04

```

*****
*           Node 9: X8           *
*           N: 1915             *
*****
*****
* Node 10 *           = Terminal Node 10 =
* N: 1826 *           = N : 89 =
*****
*****

```

Node 9 was split on X8
A case goes left if X8 <= 7.500
Improvement = 0.004 Complexity Threshold = 0.007

Node	Cases	Wgt Counts	Cost	Class
9	1915	1915.00	0.398	1
10	1826	1826.00	0.375	1
-10	89	89.00	0.155	2

Weighted Counts			
Class	Top	Left	Right
1	1198.00	1183.00	15.00
2	717.00	643.00	74.00
Within Node Probabilities			
Class	Top	Left	Right
1	0.602	0.625	0.155
2	0.398	0.375	0.845

Competitor	Split	Improve.
1 X6	2	0.002
2 X2	65.500	.725441E-03
3 X4	1	.677824E-03

```

4 X7          2          .170766E-03
5 X5          2,3        .202879E-04

```

```

*****
*           Node 10: X6          *
*           N: 1826             *
*****
=====
= Terminal Node 6 =             * Terminal Node 11 *
= N : 1111         =             * N : 715           *
=====
*****

```

```

Node 10 was split on X6
A case goes left if X6 = (2)
Improvement = 0.002      Complexity Threshold = 0.002

```

Node	Cases	Wgt	Counts	Cost	Class
10	1826	1826.00		0.375	1
-6	1111	1111.00		0.312	1
11	715	715.00		0.472	1

Weighted Counts

Class	Top	Left	Right
1	1183.00	788.00	395.00
2	643.00	323.00	320.00

Within Node Probabilities

Class	Top	Left	Right
1	0.625	0.688	0.528
2	0.375	0.312	0.472

Surrogate	Split	Assoc.	Improve.
1 X4	s 1	0.029	.887138E-03

Competitor	Split	Improve.
1 X4	1	.887138E-03
2 X2	65.500	.643112E-03
3 X7	2	.187044E-03
4 X8	6.500	.765920E-05
5 X5	2,3	.249069E-05

```

*****
*           Node 11: X4           *
*           N: 715                *
*****
*****
* Node 12 *           = Terminal Node 9 =
* N: 500 *           = N : 215       =
*****
*****
Node 11 was split on X4
A case goes left if X4 = (1)
Improvement =.726025E-03 Complexity Threshold = 0.004
Node      Cases   Wgt Counts      Cost Class
  11      715     715.00         0.472 1
  12      500     500.00         0.426 1
  -9      215     215.00         0.422 2

                          Weighted Counts
Class      Top      Left      Right
  1         395.00   299.00    96.00
  2         320.00   201.00   119.00

                          Within Node Probabilities
Class      Top      Left      Right
  1         0.528    0.574    0.422
  2         0.472    0.426    0.578

Competitor      Split      Improve.
1 X7             2          .454729E-03
2 X2             65.500     .230841E-03
3 X3             1          .175819E-03
4 X5             3          .147776E-03
5 X8             3.500      .121340E-03

*****
*           Node 12: X2           *
*           N: 500                *
*****
*****
= Terminal Node 7 =           = Terminal Node 8 =
= N : 457         =           = N : 43         =
*****
*****
Node 12 was split on X2
A case goes left if X2 <= 68.500
Improvement = .319716E-03 Complexity Threshold =
.957771E-03

```

Node	Cases	Wgt Counts	Cost	Class
12	500	500.00	0.426	1
-7	457	457.00	0.409	1
-8	43	43.00	0.394	2

Weighted Counts

Class	Top	Left	Right
1	299.00	281.00	18.00
2	201.00	176.00	25.00

Within Node Probabilities

Class	Top	Left	Right
1	0.574	0.591	0.394
2	0.426	0.409	0.606

Competitor	Split	Improve.
1 X3	1	.161757E-03
2 X5	2,3	.145383E-03
3 X8	3.500	.144400E-03
4 X7	2	.782796E-

* Node 13: X8 *

* N: 2547 *

* Node 14 * = Terminal Node 19 =

* N: 2343 * = N : 204 =

Node 13 was split on X8

A case goes left if X8 <= 13.500

Improvement = 0.005 Complexity Threshold = 0.009

Node	Cases	Wgt Counts	Cost	Class
13	2547	2547.00	0.449	2
14	2343	2343.00	0.479	2
-19	204	204.00	0.117	2

Weighted Counts

Class	Top	Left	Right
1	1206.00	1180.00	26.00
2	1341.00	1163.00	178.00

Within Node Probabilities

Class	Top	Left	Right
1	0.449	0.479	0.117
2	0.551	0.521	0.883

Competitor	Split	Improve.
1 X1	1	0.004
2 X2	76.500	0.004
3 X6	2	0.003
4 X5	1,2	0.003
5 X4	1,2	0.001

```

*****
*           Node 14: X1           *
*           N: 2343              *
*****
*****
* Node 15 *           = Terminal Node 18 =
* N: 1223 *           = N : 1120       =
*****

```

Node 14 was split on X1

A case goes left if X1 = (1)

Improvement = 0.004 Complexity Threshold = 0.018

Node	Cases	Wgt Counts	Cost	Class
14	2343	2343.00	0.479	2
15	1223	1223.00	0.430	1
-18	1120	1120.00	0.381	2

Weighted Counts

Class	Top	Left	Right
1	1180.00	727.00	453.00
2	1163.00	496.00	667.00

Within Node Probabilities

Class	Top	Left	Right
1	0.479	0.570	0.381
2	0.521	0.430	0.619

Surrogate	Split	Assoc.	Improve.
1 X5	s 2	0.520	0.003
2 X7	s 1	0.359	.575521E-04
3 X3	s 1		
4 X4	s 2,3,4,5	0.205	.518080E-03

Competitor	Split	Improve.
1 X6	2	0.003
2 X2	76.500	0.003
3 X5	1,2	0.003
4 X4	1,2	0.002
5 X8	2.500	.894688E-03

```

*****
*           Node 15: X6           *
*           N: 1223              *
*****
*****
* Node 16 *           * Node 17 *
* N: 685  *           * N: 538  *
*****
Node 15 was split on X6
A case goes left if X6 = (2)
Improvement = 0.003      Complexity Threshold = 0.006
  Node   Cases   Wgt Counts   Cost   Class
    15    1223    1223.00    0.430  1
    16     685     685.00    0.333  1
    17     538     538.00    0.449  2

                          Weighted Counts
Class      Top      Left      Right
  1         727.00   472.00   255.00
  2         496.00   213.00   283.00

                          Within Node Probabilities
Class      Top      Left      Right
  1         0.570   0.667   0.449
  2         0.430   0.333   0.551

Surrogate      Split      Assoc.      Improve.
1 X4           s 1         0.235      0.001
2 X7           s 2         0.079      .191938E-03
3 X5           s 1,2       0.029      .502828E-03

Competitor      Split      Improve.
1 X2           77.500      0.003
2 X4           1,2        0.002
3 X8           6.500       .725299E-03
4 X5           1,2        .502828E-03
5 X7           2          .191938E-03

*****
*           Node 16: X2           *
*           N: 685              *
*****
=====
= Terminal Node 12 =           = Terminal Node 13 =
= N : 515           =           = N : 170           =
=====

```

Node 16 was split on X2

A case goes left if X2 <= 78.500

Improvement = 0.002 Complexity Threshold = 0.001

Node	Cases	Wgt Counts	Cost	Class
16	685	685.00	0.333	1
-12	515	515.00	0.266	1
-13	170	170.00	0.469	2

Weighted Counts

Class	Top	Left	Right
1	472.00	388.00	84.00
2	213.00	127.00	86.00

Within Node Probabilities

Class	Top	Left	Right
1	0.667	0.734	0.469
2	0.333	0.266	0.531

Competitor	Split	Improve.
1 X5	1,2	.333717E-03
2 X8	7.500	.301693E-03
3 X4	2,5	.119021E-03
4 X7	1	.773065E-04

```
*           Node 17: X2           *
*           N: 538                 *
```

```
*****
* Node 18 *           = Terminal Node 17 =
* N: 379 *           = N : 159           =
*****
*****
```

Node 17 was split on X2

A case goes left if X2 <= 77.500

Improvement = 0.001 Complexity Threshold = 0.003

Node	Cases	Wgt Counts	Cost	Class
17	538	538.00	0.449	2
18	379	379.00	0.484	1
-17	159	159.00	0.293	2

Weighted Counts

Class	Top	Left	Right
1	255.00	205.00	50.00
2	283.00	174.00	109.00

Within Node Probabilities

Class	Top	Left	Right
1	0.449	0.516	0.293

2	0.551	0.484	0.707
---	-------	-------	-------

Competitor		Split	Improve.
1 X4	1,2		0.001
2 X8		6.500	.562342E-03
3 X7	2		.188563E-03
4 X5	1		.604805E-04

```

*****
*           Node 18: X4           *
*           N: 379                 *
*****
*****
* Node 189*                       = Terminal Node 16 =
* N: 256 *                         = N : 123           =
*****
*****

```

Node 18 was split on X4

A case goes left if X4 = (1,2)

Improvement = 0.001 Complexity Threshold = 0.004

Node	Cases	Wgt	Counts	Cost	Class
18	379		379.00	0.484	1
19	256		256.00	0.395	1
-16	123		123.00	0.335	2

Weighted Counts

Class	Top	Left	Right
1	205.00	161.00	44.00
2	174.00	95.00	79.00

Within Node Probabilities

Class	Top	Left	Right
1	0.516	0.605	0.335
2	0.484	0.395	0.665

Competitor		Split	Improve.
1 X8		6.500	.636006E-03
2 X7	2		.171436E-03
3 X2		74.500	.118889E-03
4 X5	1,2		.189053E-05

```

*****
*           Node 19: X8           *
*           N: 256                *
*****
=====
= Terminal Node 14 =                = Terminal Node 15 =
= N : 235          =                = N : 21          =
=====
Node 19 was split on X8
A case goes left if X8 <= 6.500
Improvement = .730727E-03 Complexity Threshold = 0.001
  Node      Cases   Wgt Counts      Cost Class
    19      256     256.00         0.395 1
   -14      235     235.00         0.359 1
   -15       21      21.00         0.220 2

                          Weighted Counts
Class      Top      Left      Right
  1         161.00   156.00    5.00
  2          95.00   79.00    16.00

                          Within Node Probabilities
Class      Top      Left      Right
  1         0.605   0.641    0.220
  2         0.395   0.359    0.780

Competitor      Split      Improve.
  1 X2           71.500     .117748E-03
  2 X4           1         .801968E-04
  3 X5           1,2       .320761E-04
  4 X7           2         .468840E-05

*****
*           Node 20: X2           *
*           N: 2421              *
*****
*****
* Node 21 *                = Terminal Node 24 =
* N: 1274 *                = N : 1147          =
*****
Node 20 was split on X2
A case goes left if X2 <= 72.500
Improvement = 0.005 Complexity Threshold = .968840E-
03

```

Node	Cases	Wgt	Counts	Cost	Class
20	2421		2421.00	0.219	2
21	1274		1274.00	0.317	2
-24	1147		1147.00	0.113	2
Weighted Counts					
Class	Top		Left		Right
1	574.00		432.00		142.00
2	1847.00		842.00		1005.00
Within Node Probabilities					
Class	Top		Left		Right
1	0.219		0.317		0.113
2	0.781		0.683		0.887
Surrogate		Split	Assoc.	Improve.	
1 X8	s	2.500	0.120	0.002	
2 X7	s 1		0.070	0.001	
Competitor		Split		Improve.	
1 X1	1			0.002	
2 X8		2.500		0.002	
3 X5	2			0.002	
4 X7	1			0.001	
5 X6	2			.600538E-03	

```

*****
*           Node 21: X1           *
*           N: 1274              *
*****
*****
* Node 22 *           = Terminal Node 23 =
* N: 252 *           = N : 1022         =
*****
*****
Node 21 was split on X1
A case goes left if X1 = (1)
Improvement = 0.002      Complexity Threshold = 0.001

```

Node	Cases	Wgt	Counts	Cost	Class
21	1274		1274.00	0.317	2
22	252		252.00	0.487	2
-23	1022		1022.00	0.276	2
Weighted Counts					
Class	Top		Left		Right
1	432.00		129.00		303.00
2	842.00		123.00		719.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.317	0.487	0.276
2	0.683	0.513	0.724

Competitor	Split	Improve.
1 X8	5.500	0.001
2 X5	2	0.001
3 X2	69.500	.812992E-03
4 X6	2	.774253E-03
5 X7	1	.391773E-03

```

*****
*           Node 22: X8           *
*           N: 252               *
*****
*****
* Node 23 *           = Terminal Node 22 =
* N: 211 *           = N : 41           =
*****
*****

```

Node 22 was split on X8
A case goes left if X8 <= 3.500
Improvement = 0.001 Complexity Threshold = 0.002

Node	Cases	Wgt Counts	Cost	Class
22	252	252.00	0.487	2
23	211	211.00	0.451	1
-22	41	41.00	0.180	2

Weighted Counts			
Class	Top	Left	Right
1	129.00	121.00	8.00
2	123.00	90.00	33.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.487	0.549	0.180
2	0.513	0.451	0.820

Competitor	Split	Improve.
1 X2	69.500	.805961E-03
X5	2	.372129E-03
3 X6	2	.231821E-03
4 X4	2,3	.203113E-03
5 X7	1	.787043E-04

```

*****
*           Node 23: X2           *
*           N: 211                *
*****
=====
= Terminal Node 20 =           = Terminal Node 21 =
= N : 140          =           = N : 71          =
=====
Node 23 was split on X2
A case goes left if X2 <= 69.500
Improvement = .615736E-03 Complexity Threshold = 0.002
  Node      Cases   Wgt Counts      Cost  Class
    23      211     211.00      0.451  1
   -20      140     140.00      0.366  1
   -21       71      71.00      0.385  2

                          Weighted Counts
Class      Top      Left      Right
  1         121.00   92.00    29.00
  2          90.00   48.00    42.00

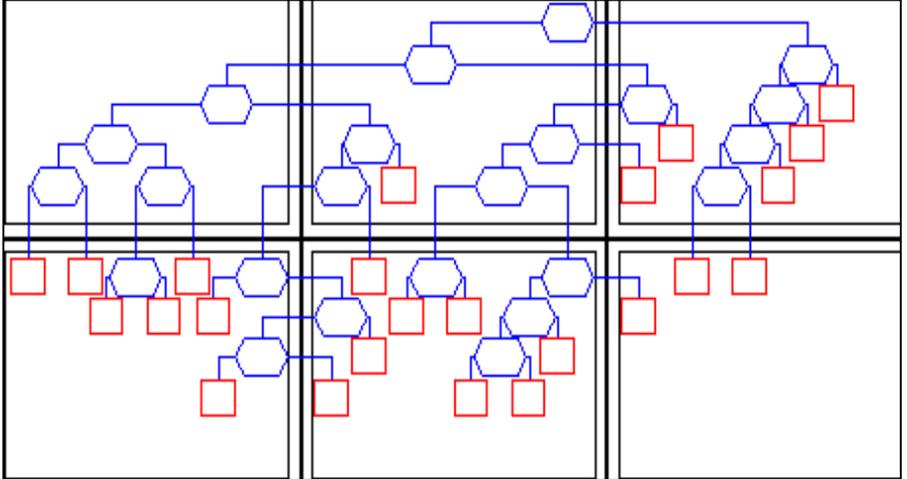
                          Within Node Probabilities
Class      Top      Left      Right
  1         0.549   0.634    0.385
  2         0.451   0.366    0.615

Surrogate      Split      Assoc.      Improve.
1 X8           s          2.500       0.016       .640750E-06

Competitor      Split      Improve.
1 X4            1,2,3,4      .201742E-03
2 X5            2          .186036E-03
3 X6            2          .101767E-03
4 X7            1          .715293E-04
5 X3            3          .303675E-04

```

LAMPIRAN D
Konstruksi Pohon Klasifikasi Optimal



LAMPIRAN E

Program R untuk Pembagian Data *Learning* dan *Testing*

```

datata<-read.csv("D:/run.csv",header=TRUE) #load the data

#Split the data frame
splitDataFrame<-function(dataframe,seed=null,n=learnSize){
  if(!is.null(seed))set.seed(seed)
  index<-1:nrow(dataframe)
  learnindex<-sample(index,n)
  learnset<-dataframe[learnindex,]
  testset<-dataframe[-learnindex,]
  list(learnset=learnset,testset=testset)
}

# Learning Data 75% and Testing Data 25%
split<-splitDataFrame(datata,NULL,round(nrow(datata)*0.75))
learn75<-split$learnset
test25<-split$testset
write.csv(learn75, "D:\\learn75.csv")
write.csv(test25, "D:\\test25.csv")

# Learning Data 80% and Testing Data 20%
split<-splitDataFrame(datata,NULL,round(nrow(datata)*0.80))
learn80<-split$learnset
test20<-split$testset
write.csv(learn80, "D:\\learn80.csv")
write.csv(test20, "D:\\test20.csv")

# Learning Data 85% and Testing Data 15%
split<-splitDataFrame(datata,NULL,round(nrow(datata)*0.85))
learn85<-split$learnset
test15<-split$testset
write.csv(learn85, "D:\\learn85.csv")
write.csv(test15, "D:\\test15.csv")

# Learning Data 90% and Testing Data 10%
split<-splitDataFrame(datata,NULL,round(nrow(datata)*0.90))
learn90<-split$learnset
test10<-split$testset

```

```
write.csv(learn90, "D:\\learn90.csv")
write.csv(test10, "D:\\test10.csv")

# Learning Data 95% and Testing Data 5%
split<-splitDataFrame(datata,NULL,round(nrow(datata)*0.95))
learn95<-split$learnset
test5<-split$testset
write.csv(learn95, "D:\\learn95.csv")
write.csv(test5, "D:\\test5.csv")

# source : (Turner, 2011)
```

DAFTAR PUSTAKA

- Andriyashin, A. 2005. *Financial Application of Classification and Regression Trees*. Berlin: Master Thesis, CASE-Center of Applied Statistics and Economics Humboldt University.
- Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc. New Jersey.
- Badan Pusat Statistik Indonesia (BPS). 2011. *Hasil Survei Sosial Ekonomi Nasional Tahun 2011 Provinsi Jawa Timur*. Jakarta: BPS.
- Badan Pusat Statistik Indonesia (BPS). 2011. *Statistik Penduduk Lanjut Usia Tahun 2011*. Jakarta: BPS
- Badan Pusat Statistik Jawa Timur (BPS). 2011. *Profil Penduduk Lanjut Usia Jawa Timur*. Surabaya : BPS
- Breiman, L. Friedman, J. H. Olshen, R. A. and Stone, C. K. 1993. *Classification And Regression Tree*. New York, NY: Chapman And Hall
- Depkes RI., 1992. *Undang-Undang Kesehatan (UU RI No. 23 Tahun 1992 Tentang Kesehatan)*. Indonesian Legal Center Publishing. Jakarta
- Cristie, D. 2012. *Faktor-Faktor yang Berpengaruh Terhadap Aktivitas Ekonomi Lansia di Provinsi Jawa Timur dengan Regresi Logistik Biner*. Surabaya: Tugas Akhir Program Studi Sarjana Fakultas MIPA Institut Teknologi Sepuluh Nopember.
- Hurlock, E. B. 2000. *Psikologi Perkembangan Suatu Pendekatan Sepanjang Rentang Kehidupan (edisi kelima)*. Jakarta: Erlangga
- Johnson, R. A. dan Winchern, D. W.(2007). *Applied Multivariate Statistical Analysis, 5th Edition*. New Jersey: Prentice Hall, Englewood Clifft.
- Kurniawan, A. 2010. *Belajar Mudah SPSS Untuk Pemula*. Yogyakarta: Media Kom
- Lewis, R. J. 2000. *An Introduction to Classification and Regression Trees (CART) Analysis*. Annual Meeting of the Society for Academic Emergency Medicine. California, UCLA Medical Center.

- Pratiwi, F.E. 2014. *Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification And Regression Trees Di Provinsi Sulawesi Utara*. Surabaya: Tugas Akhir Program Studi Sarjana Fakultas MIPA Institut Teknologi Sepuluh Nopember.
- Wahyuningsih, M. 2011. *detik helath*.
<http://health.detik.com/read/2011/12/06/170435/1784303/763/ini-dia-5-provinsi-dengan-jumlah-lansia-paling-banyak> (diakses 23 Februari 2015 pukul 17.00)
- Walpole, R.E dan Mayer, R.H. 1995. *Ilmu Peluang dan Statistik Untuk Insinyur dan Ilmuwan* (terjemahan RK Sembiring). Terjemahan Keempat. Bandung: ITB.

BIODATA PENULIS



Penulis dengan nama lengkap Pitri Ariska Susilowati dilahirkan di Tulungagung pada tanggal 2 April 1992 yang merupakan anak pertama dan terakhir. Riwayat pendidikan penulis ditempuh di SDN Karangtalun 2, SMP Negeri 1 Kalidawir dan SMA Negeri 1 Boyolangu (SMABOY), Diploma III ITS pada tahun 2010. Tahun 2013 penulis melanjutkan jenjang Sarjana melalui program Lintas Jalur di Jurusan Statistika ITS.

Penulis pernah memiliki pengalaman kerja sebagai *Caller* (Call Back) dan entrier di *Enciety Bisnis Consult* pada tahun 2013 sampai dengan 2014. Segala saran dan kritik yang membangun serta apabila pembaca ingin berdiskusi lebih lanjut mengenai Tugas Akhir ini atau materi lain yang berhubungan dengan penulis, dapat mengirimkan email ke ariskasusilowati@gmail.com