TUGAS AKHIR – KS184822

KLASIFIKASI BALITA TERKAIT KASUS *STUNTING* DI KOTA SURABAYA MENGGUNAKAN *CLASSIFICATION AND REGRESSION TREES* DENGAN *ADAPTIVE SYNTHETIC NOMINAL* DAN *ADAPTIVE SYNTHETIC K-NEAREST NEIGHBOR*

WINANTRI NURBUWAH INSYROH
NRP 062116 4000 0034

Dosen Pembimbing
Irhamah, M.Si, Ph.D
Erma Oktania Permatasari, S.Si., M.Si

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS SAINS DAN ANALITIKA DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2020

TUGAS AKHIR – KS184822

KLASIFIKASI BALITA TERKAIT KASUS *STUNTING* DI KOTA SURABAYA MENGGUNAKAN *CLASSIFICATION AND REGRESSION TREES* DENGAN *ADAPTIVE SYNTHETIC NOMINAL* DAN *ADAPTIVE SYNTHETIC K-NEAREST NEIGHBOR*

WINANTRI NURBUWAH INSYROH
NRP 062116 4000 0034

Dosen Pembimbing
Irhamah, M.Si, Ph.D
Erma Oktania Permatasari, S.Si., M.Si

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS SAINS DAN ANALITIKA DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2020

FINAL PROJECT – KS184822

CLASSIFICATION OF STUNTED CHILD CASES IN SURABAYA USING CLASSIFICATION AND REGRESSION TREES WITH ADAPTIVE SYNTHETIC NOMINAL DAN ADAPTIVE SYNTHETIC K-NEAREST NEIGHBOR

WINANTRI NURBUWAH INSYROH
SN 062116 4000 0034

Supervisors
Irhamah, M.Si, Ph.D
Erma Oktania Permatasari, S.Si., M.Si

UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF SCIENCE AND DATA ANALYTICS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2020

LEMBAR PENGESAHAN

KLASIFIKASI BALITA TERKAIT KASUS *STUNTING* DI KOTA SURABAYA MENGGUNAKAN CLASSIFICATION AND REGRESSION TREES DENGAN ADAPTIVE SYNTHETIC NOMINAL DAN ADAPTIVE SYNTHETIC K-NEAREST NEIGHBOR

TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Statistika
pada
Program Studi Sarjana Departemen Statistika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember

Oleh :
Winantri Nurbuwah Insyroh
NRP. 062116 4000 0034

Disetujui oleh Pembimbing Tugas Akhir:

Irhamah, M.Si, Ph.D
NIP. 19780406 200112 2 002 ( )

Erma Oktania Permatasari, S.Si., M.Si
NIP. 19881007 201404 2 002 ( )

Mengetahui,
Kepala Departemen Statistika

Dr. Dra. Kartika Fithriasari, M.Si
NIP. 19691212 199303 2 002

SURABAYA, AGUSTUS 2020

(Halaman ini sengaja dikosongkan)

KLASIFIKASI BALITA TERKAIT KASUS *STUNTING* DI KOTA SURABAYA MENGGUNAKAN *CLASSIFICATION AND REGRESSION TREES* DENGAN *ADAPTIVE SYNTHETIC NOMINAL* DAN *ADAPTIVE SYNTHETIC K-NEAREST NEIGHBOR*

Identitas Mahasiswa	**: Winantri Nurbuwah Insyroh**
NRP	**: 062116 4000 0034**
Departemen	**: Statistika**
Dosen Pembimbing	**: Irhamah, M.Si., Ph.D**
	: Erma Oktania Permatasari, S.Si., M.Si

Abstrak

Stunting merupakan masalah gizi kronis yang disebabkan oleh defisiensi nutrien selama seribu hari pertama kehidupan yang menyebabkan penurunan kemampuan kognitif dan motorik serta penurunan performa kinerja fisik. Jumlah balita stunting di Indonesia kini mencapai 27,67 persen. Kemudian di Jawa Timur, pada tahun 2019, terdapat 32,8 persen kasus balita stunting. Sedangkan di Surabaya, angka stunting hampir mendekati 30 persen. Untuk mengurangi angka stunting, perlu diketahui faktor apa saja yang dapat mengindikasikan terjadinya stunting pada balita. Pada penelitian ini dilakukan klasifikasi terhadap faktor-faktor yang diduga menyebabkan kasus stunting pada balita di Kota Surabaya menggunakan metode klasifikasi CART. Data merupakan data sekunder yang diperoleh dari Dinas Kesehatan Kota Surabaya. Data pada penelitian ini merupakan data imbalance sehingga dilakukan penanganan data imbalance dengan ADASYN-N dan ADASYN-kNN. Kesimpulan yang dihasilkan pada penelitian ini adalah metode ADASYN-N memiliki nilai AUC lebih baik dibanding metode ADASYN-kNN, sehingga hasil dari klasifikasi faktor-faktor yang mengindikasikan terjadinya stunting pada balita di Kota Surabaya berdasarkan metode terbaik adalah tinggi badan balita dibandingkan umur, ketersediaan jamban, dan tidak adanya anggota keluarga yang merokok di dalam rumah.

Kata kunci: ADASYN-kNN, ADASYN-N, Balita, CART, Stunting

(Halaman ini sengaja dikosongkan)

CLASSIFICATION OF STUNTED CHILD CASES IN SURABAYA USING *CLASSIFICATION AND REGRESSION TREES* WITH *ADAPTIVE SYNTHETIC NOMINAL* AND *ADAPTIVE SYNTHETIC K-NEAREST NEIGHBOR*

Name	**:**	**Winantri Nurbuwah Insyroh**
Student Number	**:**	**062116 4000 0034**
Department	**:**	**Statistics**
Supervisors	**:**	**Irhamah, M.Si., Ph.D**
	:	**Erma Oktania Permatasari, S.Si., M.Si**

Abstract

Stunting is a chronic nutritional problem caused by nutrient deficiency during the first thousand days after birth which causes a decrease in cognitive and motoric skills and decreased physical performance. The number of toddlers who suffer from stunting in Indonesia has now reached 27,67%. Then in East Java, in 2019, there were 32,8% of the similar cases. Whereas in Surabaya, the stunting rate come near to 30%. To reduce those type of cases, it is necessary to know what kind factors that can be used to predict stunting in toddlers. In this study the classification of the factors suspected to predict stunting cases in toddlers in Surabaya was carried out using the CART classification method. Data that used in this study are the secondary data obtained from the Surabaya City Health Office. The data in this study are imbalance data so that the handling of imbalance data is carried out with ADASYN-N and ADASYN-kNN. The conclusion generated in this study is the ADASYN-N method has a better AUC value than the ADASYN-kNN method, so the result of the classification of factors that indicate stunting in toddlers in Surabaya based on the best method are the height compared to age, availability of latrines, and absence of family members who smoke inside the house.

Keywords: ADASYN-kNN, ADASYN-N, CART, Stunting, Toddler

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT atas rahmat dan hidayah yang diberikan sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul "Klasifikasi Balita Terkait Kasus *Stunting* di Kota Surabaya Menggunakan *Classification and Regression Trees* dengan *Adaptive Synthetic Nominal* dan *Adaptive Synthetic K-Nearest Neighbor*" dengan lancar.

Penulis menyadari bahwa Tugas Akhir ini dapat terselesaikan tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Kedua orang tua tercinta, atas segala do'a, nasehat, dan dukungan yang telah diberikan kepada penulis selama ini.
2. Dr. Dra. Kartika Fithriasari, M.Si selaku Kepala Departemen Statistika dan Dr. Santi Wulan Purnami, S.Si., M.Si selaku Sekretaris Departemen Bidang Akademik yang telah memberikan fasilitas, sarana, dan prasarana.
3. Prof. Dr. Drs. I Nyoman Budiantara, M.Si selaku dosen wali yang telah memberikan saran, dukungan, arahan, nasihat selama proses belajar di Departemen Statistika ITS.
4. Irhamah, M.Si., Ph.D dan Erma Oktania Permatasari, S.Si., M.Si selaku dosen pembimbing Tugas Akhir yang telah meluangkan waktu dan dengan sangat sabar membimbing, memberi saran, serta dukungan selama penyusunan Tugas Akhir.
5. Prof. Drs. NUR Iriawan, MIKom., Ph.D dan Adatul Mukarromah, S.Si., M.Si selaku dosen penguji yang selalu sabar dalam mengomentari serta memberikan masukan dan saran dalam penyelesaian Tugas Akhir.
6. Seluruh dosen Statistika ITS yang telah memberikan ilmu dan pengetahuan yang tak ternilai harganya, serta segenap karyawan Departemen Statistika ITS, khususnya Bapak

Pendi, Bapak Ipul, dan Bapak Umam yang selalu siap membantu dalam administrasi.

7. Sahabat Bunga Matahari tercinta : Kinan (pembimbing ketiga penulis), Cahya, Fransiska, Erika, Riris, dan Thalia yang selalu ada untuk membantu, menjadi pendengar, dan selalu memberikan dukungan kepada penulis.
8. Teman-teman Statistika ITS ∑27 Tr16ger angkatan 2016, yang selalu memberikan dukungan kepada penulis selama ini.
9. Semua teman, relasi dan berbagai pihak yang tidak bisa penulis sebutkan identitasnya satu persatu yang telah membantu dalam penulisan laporan ini.

Besar harapan penulis untuk mendapatkan kritik dan saran yang membangun sehingga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang terkait.

Surabaya, Juli 2020

Penulis

DAFTAR ISI

Halaman

LEMBAR PENGESAHAN ..... **iii**

Abstrak ..... **vii**

Abstract ..... **ix**

KATA PENGANTAR ..... **xi**

DAFTAR ISI ..... **xiii**

DAFTAR GAMBAR ..... **xv**

DAFTAR TABEL ..... **xvii**

DAFTAR LAMPIRAN ..... **xix**

BAB I PENDAHULUAN **1**

1.1 Latar Belakang 1
1.2 Rumusan Masalah 7
1.3 Tujuan Penelitian 8
1.4 Manfaat Penelitian 8
1.5 Batasan Masalah 8

BAB II TINJAUAN PUSTAKA **9**

2.1 Klasifikasi 9
2.2 CART (Classification and Regression Trees) 10
2.2.1 *Tree Building Process* 12
2.2.2 Pemangkasan Pohon Klasifikasi 15
2.2.3 Penentuan Pohon Klasifikasi Optimum 16
2.3 *Value Difference Metric* (VDM) 17
2.4 *Adaptive Synthetic Nominal* (ADASYN-N) 18
2.5 *Adaptive Synthetic*-kNN (ADASYN-kNN) 20
2.6 Ketepatan Klasifikasi 21
2.7 Balita *Stunting* 22

BAB III METODOLOGI PENELITIAN **29**

3.1 Sumber Data 29

3.2 Variabel Penelitian dan Definisi Operasional 29
3.3 Struktur Data 33
3.4 Langkah Analisis 34
3.5 Diagram Alir 35
BAB IV ANALISIS DAN PEMBAHASAN 37
4.1 *Preprocessing Data* 37
4.2 Eksplorasi Data 39
4.3 Analisis Klasifikasi Balita Terkait Kasus Stunting Menggunakan CART dengan ADASYN-N 45
4.4 Analisis Klasifikasi Balita Terkait Kasus *Stunting* Menggunakan CART dengan ADASYN-kNN 55
4.5 Perbandingan Metode 57
BAB V KESIMPULAN DAN SARAN 59
5.1 Kesimpulan 59
5.2 Saran 59
DAFTAR PUSTAKA 61
LAMPIRAN 65
BIODATA PENULIS 77

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Pohon Klasifikasi CART 11
Gambar 2. 2 WHO *Conceptual Framework of Stunting* 24
Gambar 3. 1 Diagram Alir 36
Gambar 4. 1 *Output Missing Value* 38
Gambar 4. 2 Persentase Balita Terkait Kasus Stunting 39
Gambar 4. 3 *BarChart* Balita *Stunting* di Setiap Kecamatan di Kota Surabaya 40
Gambar 4. 4 Persentase Kategori Balita Stunting Berdasarkan TB/U 40
Gambar 4. 5 *Bar Chart* Kategori Balita *Stunting* Berdasarkan (a) BB/TB dan (b) BB/U 41
Gambar 4. 6 Persentase Balita *Stunting* Makan Buah dan Sayur Setiap Hari..... 42
Gambar 4. 7 Persentase Balita Stunting yang Memperoleh PMT 43
Gambar 4. 8 Persentase Pendampingan Balita Stunting..... 44
Gambar 4. 9 Bar Chart Kelengkapan Imunisasi Balita Stunting44
Gambar 4. 10 Bar Chart Keluarga Balita Stunting yang Memilliki Sarana Air Bersih 45
Gambar 4. 11 *Complexity Parameter* terhadap Data *Imbalance*50
Gambar 4. 12 Pohon Klasifikasi Partisi 1..... 51
Gambar 4. 13 Pohon Klasifikasi Terbaik dengan ADASYN-N 53
Gambar 4. 14 Pohon Klasifikasi Terbaik dengan ADASYN-kNN 56

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

Tabel 2. 1 Ilustrasi Tabel untuk Perhitungan Proporsi Pengamatan Simpul 14
Tabel 2. 2 Ketepatan Klasifikasi 21
Tabel 2. 3 Kriteria Nilai AUC..... 22
Tabel 2. 4 Pengertian Kategori Status Gizi Balita 25
Tabel 3. 1 Variabel Penelitian..... 29
Tabel 3. 2 Struktur Data Penelitian 33
Tabel 4. 1 Pemilahan pada Simpul Anggota Keluarga Merokok49
Tabel 4. 2 *Confusion Matrix Adasyn-N* Partisi 1 51
Tabel 4. 3 Nilai AUC Hasil Klasifikasi CART dengan ADASYN-N 52
Tabel 4. 4 *Gains Nodes* Balita Mengalami Kasus *Stunting* (ADASYN-N) 53
Tabel 4. 5 Nilai AUC Hasil Klasifikasi CART dengan ADASYN-kNN 56
Tabel 4. 6 *Gains Nodes* Balita Mengalami Kasus *Stunting* (ADASYN-kNN)..... 57
Tabel 4. 7 Perbandingan Nilai AUC *Testing* pada Klasifikasi CART..... 58

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian 65
Lampiran 2. Pohon Klasifikasi CART dengan ADASYN-N ... 66
Lampiran 3. Pohon Klasifikasi CART dengan ADASYN-kNN67
Lampiran 4. *Syntax* R CART dengan ADASYN-N 69
Lampiran 5. *Syntax* R CART dengan ADASYN-kNN 73
Lampiran 6. Surat Keterangan 76

(Halaman ini sengaja dikosongkan)

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan pada bayi merupakan pertumbuhan cepat (*growth spurt*) yang dimulai pada usia 3 bulan hingga usia 2 tahun, kemudian pertumbuhan terus berlanjut walaupun sedikit melambat pada usia 2 tahun sampai usia anak 5 tahun dan akan memengaruhi kemampuan motoris, sosial, emosional, dan perkembangan kognitif (Astari, 2006). Masa balita merupakan masa penting dalam perkembangan dan pertumbuhan manusia karena menjadi penentu keberhasilan pertumbuhan serta perkembangan anak di masa mendatang. Masa tumbuh kembang di usia ini merupakan *golden age* atau masa keemasan karena merupakan masa penting untuk mengoptimalkan pertumbuhan dan perkembangan anak. Masa balita juga merupakan masa yang paling rawan terhadap kekurangan energi dan protein sehingga asupan gizi yang baik sangat diperlukan dalam proses pertumbuhan dan perkembangan karena apabila zat gizi tubuh tidak terpenuhi maka dapat menyebabkan dampak yang serius bagi anak (Waryana, 2010). Pemenuhan gizi terhadap anak sebaiknya dimulai sejak dalam kandungan. Salah satu permasalahan yang disebabkan oleh kekurangan gizi kronis sejak dalam masa kandungan adalah *stunting*.

Stunting adalah kondisi ketika balita memiliki panjang atau tinggi badan yang kurang jika dibandingkan dengan umur. Kondisi ini diukur dengan panjang atau tinggi badan yang lebih dari minus dua standar deviasi median standar pertumbuhan anak dari WHO. *Stunting* dikelompokkan menjadi pendek dan sangat pendek. Pendek atau *stunted* adalah status gizi balita TB/U yang memiliki *z score* antara -3 SD hingga –2SD, sedangkan sangat pendek atau *severely stunted* adalah status gizi balita TB/U yang nilainya kurang dari -3SD (Kementerian Kesehatan, 2010). *Stunting* merupakan masalah gizi kronis yang disebabkan oleh defisiensi nutrien selama seribu hari pertama kehidupan. Hal ini

dapat menimbulkan gangguan perkembangan fisik anak yang *irreversible*, sehingga menyebabkan penurunan kemampuan kognitif dan motorik serta penurunan performa kinerja fisik. Anak *stunting* memiliki rata-rata skor *Intelligence Quotient* (IQ) sebelas poin lebih rendah dibandingkan rata-rata skor IQ pada anak normal. Gangguan tumbuh kembang pada anak akibat kekurangan gizi kronis selama seribu hari pertama kehidupan ini bila tidak mendapatkan intervensi sejak dini akan berlanjut hingga dewasa (Trihono, 2015). Kondisi yang berlanjut hingga dewasa ini juga dapat menyebabkan kesulitan dalam kemampuan kognitif dan prestasi belajar, dan beresiko tinggi mengalami berbagai penyakit seperti diabetes, penyakit jantung dan pembuluh darah, kanker, stroke, dan disabilitas pada usia tua. *Stunting* terjadi mulai dari dalam kandungan dan baru terlihat saat anak berusia dua tahun sehingga pencegahan sejak dini harus dilakukan. Permasalahan *stunting* ini juga merupakan permasalahan yang cukup serius di Indonesia.

Data prevalensi balita *stunting* yang dikumpulkan WHO, Indonesia termasuk ke dalam negara ketiga dengan prevalensi tertinggi di regional Asia Tenggara/South-East Asia Regional (SEAR), dengan rata-rata prevalensi balita *stunting* di Indonesia tahun 2005-2017 adalah 36,4 persen. Angka *stunting* di Indonesia terus mengalami penurunan dari tahun 2013 hingga 2019. Meski terus mengalami penurunan, namun angka tersebut masih jauh dari target pemerintah dengan angka *stunting* yang ditargetkan di bawah 20 persen dalam lima tahun ke depan (Mustika, 2020). Gentingnya permasalahan balita *stunting* ini diperkuat oleh desakan Presiden Joko Widodo untuk melakukan percepatan penyelesaian masalah *Stunting* di acara Musyawarah Rencana Pembangunan (Musrenbang) Nasional Rancangan RPJMN 2020-2024. Pesan dari Presiden itu juga seolah mengkonfirmasi bahwa penurunan angka *stunting* di Indonesia masih belum cukup menggembirakan. Menurut hasil riset studi status gizi balita Indonesia (SSGBI) 2019, jumlah balita *stunting* di Indonesia saat ini mencapai 27,67 persen. Artinya, terdapat 6,3 juta dari populasi

23 juta balita di Indonesia yang mengidap masalah *stunting*. Jumlah yang melampaui nilai standar maksimal dari Organisasi Kesehatan Dunia (WHO) yakni sebesar 20 persen atau seperlima dari jumlah total anak balita dalam suatu negara. Menurut pernyataan Muktamar Umakaapa peneliti di Merial Institute Magister Ilmu Kebijakan dan Hukum Kesehatan di Fakultas Kesehatan Masyarakat Universitas Indonesia yang dimuat pada Detik News pada 14 Januari 2020, tingginya kasus *stunting* ini memberikan dampak yang buruk terhadap pembangunan dan kemajuan Indonesia karena mengancam produktivitas sumber daya manusia Indonesia, mengurangi daya saing sehingga menyebabkan kerugian ekonomi. Kerugian ekonomi yang terjadi akibat kasus *Stunting* menurut World Bank dapat mencapai 2 hingga 3 persen dari Produk Domestik Bruto (PDB) per tahun suatu negara.

Di Indonesia, wilayah Jawa Timur merupakan salah satu provinsi yang memiliki angka *stunting* yang cukup tinggi. Berdasarkan data Riset Kesehatan Dasar Kementrian Kesehatan tahun 2019, di Provinsi jawa Timur terdapat 32,8 persen kasus balita *Stunting* dimana angka nasional balita *stunting* sebesar 27,67 persen. Sedangkan di Surabaya yang merupakan ibukota Jawa Timur, menurut hasil Riskesdas Provinsi Jawa Timur pada tahun 2018, tercatat angka *stunting* di Kota Surabaya hampir mendekati 30 persen, hal ini masih cukup jauh dari target pemerintah untuk menurunkan angka *stunting* dibawah 20 persen sehingga penanganan terhadap balita *stunting* masih terus dilakukan oleh Pemerintah Kota Surabaya. Pada Tahun 2019 dilakukan pemeriksaan terhadap balita di setiap kelurahan di Kota Surabaya oleh Kader di Posyandu bekerjasama dengan pihak Puskesmas untuk mengetahui balita yang diduga mengalami kasus *stunting*, kemudian dari *screening* tersebut diperoleh hasil bahwa sekitar 30 persen balita di Kota Surabaya diduga mengalami kasus *stunting*. Selanjutnya pada Agustus 2019 hingga Februari 2020 dilakukan pengecekan langsung oleh tenaga kesehatan terhadap balita yang diduga mengalami kasus *stunting*

tersebut. Pengecekan dilakukan oleh tenaga kesehatan dengan menggunakan peralatan yang telah sesuai dengan standar kesehatan dan pengecekan dilakukan dari rumah ke rumah. Pengecekan pada periode tersebut dilakukan pada 4757 balita dari data yang dikumpulkan oleh Kader Posyandu. Hasil dari pengecekan menunjukkan bahwa balita yang positif mengalami kasus *stunting* di Kota Surabaya sebanyak 4614 balita dari 4757 balita yang diperiksa. Kasus balita *stunting* di Kota Surabaya yang cukup tinggi ini menunjukkan perlunya diketahui faktor-faktor apa saja yang mengindikasikan terjadinya kasus balita *stunting* sehingga penanganan yang dilakukan oleh Pemerintah Kota Surabaya dapat lebih optimal.

World Health Organization (WHO) memiliki *conceptual framework* yang berisi faktor-faktor yang menyebabkan terhambatnya pertumbuhan dan perkembangan balita (penyebab balita *stunting*). Terdapat 4 faktor utama yaitu faktor keluarga dan rumah tangga, faktor kelengkapan makanan yang tidak memadai, faktor menyusui, serta faktor infeksi. Faktor keluarga dan rumah tangga terbagi menjadi dua faktor yaitu faktor maternal dan faktor lingkungan, sedangkan faktor kelengkapan makanan yang tidak memadai terdapat tiga faktor yaitu kualitas makanan yang buruk, praktik pemberian makanan yang tidak memadai, serta keamanan dari air dan makanan. Dari uraian diatas terlihat bahwa penyebab dari balita *stunting* terdiri dari banyak faktor, setiap faktor masih memiliki beberapa variabel sehingga dapat dikatakan bahwa kasus *stunting* pada balita disebabkan oleh hal-hal yang cukup kompleks sehingga diperlukan pengklasifikasian terhadap balita untuk mengetahui faktor-faktor yang mengindikasikan terjadinya *stunting* pada balita dari banyak faktor yang ada. Terdapat beberapa metode klasifikasi yang umum digunakan antara lain CART (*Classification and Regression Trees*), SVM, Naïve Bayes, dan metode klasifikasi lainnya. Salah satu metode klasifikasi yang mudah untuk dipahami, dapat digunakan untuk data dengan skala besar serta banyak variabel adalah metode CART (*Classification and Regression Trees*). Oleh karena itu,

pada penelitian ini dilakukan klasifikasi untuk mengetahui faktor-faktor yang mengindikasikan terjadinya kasus *stunting* pada balita di Kota Surabaya dengan menggunakan metode CART. Hal ini sesuai dengan besarnya data yang digunakan pada penelitian ini karena penelitian dilakukan terhadap rumah tangga yang ada di Kota Surabaya, metode ini juga mudah dipahami sehingga diharapkan dapat mempermudah Dinas Kesehatan Kota Surabaya dalam memperoleh informasi terkait hasil analisis klasifikasi balita *stunting* ini. Metode CART (*Classification and Regression Trees*) merupakan suatu metode nonparametrik yang menggambarkan hubungan antara variabel respon (dependen) dengan satu atau lebih variabel prediktor (variabel independen) yang bertujuan untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian (Breiman, 1993).

Terdapat beberapa penelitian terkait metode CART, yaitu penelitian yang dilakukan oleh Novie Eriska Aritonang pada tahun 2016 dengan penelitian berjudul Klasifikasi Status Kerja pada Angkatan Kerja Kota Semarang Tahun 2014 Menggunakan Metode CHAID dan CART, pada penelitian tersebut hasil klasifikasi dengan menggunakan metode CART memiliki ketepatan klasifikasi lebih tinggi dibanding menggunakan metode CHAID yaitu ketepatan klasifikasi metode CART sebesar 72,79% dan 72,63% untuk metode CHAID. Selanjutnya penelitian Rokhana Dwi Bekti pada tahun 2017 yang berjudul Analisis Pengaruh Lokasi dan Karakteristik Konsumen dalam Memilih Minimarket dengan Metode Regresi Logistik dan CART, pada penelitian ini metode CART memberikan hasil ketepatan klasifikasi sebesar 90% yang mana lebih baik dibandingkan dengan hasil ketepatan klasifikasi regresi logistik sebesar 88%. Kemudian pada penelitian oleh Agung Waluyo pada tahun 2014 mengenai perbandingan klasifikasi nasabah kredit menggunakan regresi logistik biner dan CART, diperoleh kesimpulan bahwa kinerja CART lebih baik dibandingkan dengan regresi logistik biner. Pada penelitian mengenai kasus balita *Stunting*, terdapat

penelitian dari Bunga Ch Rosha mengenai analisis determinan *Stunting* anak 0-23 bulan pada daerah miskin di Jawa Tengah dan Jawa Timur dan diketahui faktor-faktor yang mempengaruhi *stunting* anak 0-23 bulan pada daerah miskin di Jawa Tengah dan Jawa Timur. Penelitian juga dilakukan oleh Eko Setiawan pada tahun 2018 untuk mengetahui faktor-faktor yang berhubungan dengan kejadian *Stunting* pada anak usia 24-59 bulan di wilayah kerja puskesmas andalas Kecamatan Padang Timur Kota Padang, pada penelitian tersebut faktor tingkat pendidikan ibu memiliki hubungan paling dominan.

Persentase balita *stunting* pada penelitian ini jauh lebih besar dibandingkan dengan persentase balita normal, perbedaan yang cukup besar pada persentase balita di kelas normal dengan persentase balita di kelas *stunting* tersebut mengindikasikan terjadinya permasalahan *imbalance* sehingga perlu dilakukan penanganan *imbalance*. Terdapat beberapa metode untuk menangani kasus *imbalance*, antara lain metode SMOTE (*Synthetic Minority Over-sampling Technique*) serta metode ADASYN (*Adaptive Synthetic*). SMOTE yang biasa digunakan untuk data numerik kemudian dikembangkan menjadi SMOTE-N untuk digunakan pada data nominal. Sedangkan untuk pengembangan ADASYN pada data nominal juga dikembangkan oleh Kurniawati pada tahun 2017 menjadi ADASYN-N dan ADASYN-kNN. ADASYN-N dan ADASYN-kNN dapat digunakan untuk menangani ketidakseimbangan data dengan kelas nominal dengan jumlah kategori pada masing-masing kelasnya adalah 2. Sedangkan pada penelitian Rahayu (2017), dilakukan analisis perbandingan metode ADASYN-N dan ADASYN-kNN untuk data dengan fitur *nominal-multi categories* dengan menggunakan *classifier* Random Forest, kemudian diperoleh kesimpulan bahwa ADASYN-N menunjukkan akurasi yang lebih baik dari ADASYN-kNN dalam mengatasi ketidakseimbangan distribusi kelas pada data dengan fitur *nominal-multi categories*. Selanjutnya pada penelitian Kurniawati, dkk (2018), diperoleh hasil bahwa metode

ADASYN-kNN memiliki ketepatan klasifikasi lebih baik dibandingkan ADASYN-N pada kasus *multiclass imbalance* pada data uji coba laboraturium, ADASYN-N juga meningkatkan akurasi lebih baik dibandingkan dengan metode SMOTE-N dengan menggunakan *classifier* Naïve Bayes.

Pada penelitian ini akan dilakukan klasifikasi terhadap balita untuk mengetahui faktor-faktor yang mengindikasikan terjadinya kasus *stunting* pada balita di Kota Surabaya menggunakan metode klasifikasi CART dengan penanganan *imbalance* ADASYN-N dan ADASYN-kNN. Metode penanganan *imbalance* untuk data nominal ini dilakukan karena akan dilakukan klasifikasi pada dua kelas yaitu kelas balita normal dan kelas balita *stunting*. Sedangkan untuk variabel prediktor yang digunakan pada penelitian ini mencakup faktor keluarga dan rumah tangga, faktor kelengkapan makanan yang tidak memadai, faktor menyusui, serta faktor infeksi. Setelah dilakukan klasifikasi kemudian akan dilakukan perbandingan metode penanganan *imbalance* terbaik berdasarkan nilai AUC tertinggi. Sehingga diharapkan didapatkan hasil klasifikasi terbaik agar dapat menjadi tambahan informasi bagi instansi sebagai upaya dalam menurunkan angka *stunting* di Kota Surabaya.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah pada penelitian ini yaitu.

1. Bagaimana karakteristik dari faktor-faktor yang diduga mengindikasikan terjadinya kasus *stunting* pada balita di Kota Surabaya?
2. Bagaimana hasil klasifikasi balita untuk mengetahui faktor-faktor yang mengindikasikan terjadinya kasus *stunting* pada balita di Kota Surabaya menggunakan metode CART (*Classification and Regression Trees*) dengan ADASYN-N dan ADASYN-kNN?
3. Bagaimana hasil perbandingan metode penanganan *imbalance* ADASYN-N dan ADASYN-kNN?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah.

1. Mendeskripsikan karakteristik faktor-faktor yang diduga mengindikasikan terjadinya kasus balita *stunting* di Kota Surabaya berdasarkan data Dinas Kesehatan Kota Surabaya.
2. Mendapatkan hasil klasifikasi balita untuk mengetahui faktor-faktor yang mengindikasikan terjadinya kasus *stunting* pada balita di Kota Surabaya menggunakan metode CART (*Classification and Regression Trees*) dengan ADASYN-N dan ADASYN-kNN.
3. Mendapatkan hasil perbandingan dari metode penanganan *imbalance* ADASYN-N dan ADASYN-kNN berdasarkan nilai AUC.

1.4 Manfaat Penelitian

Manfaat yang diharapkan pada penelitian ini yaitu dapat memberikan informasi kepada pembaca mengenai metode klasifikasi CART dengan metode penanganan *imbalance* ADASYN-N dan ADASYN-kNN. Diharapkan pula hasil dari penelitian ini dapat menjadi tambahan informasi terkait faktor-faktor yang mengindikasikan terjadinya *stunting* di Kota Surabaya bagi Dinas Kesehatan Kota Surabaya sebagai upaya dalam menurunkan angka *stunting* di Kota Surabaya.

1.5 Batasan Masalah

Batasan masalah pada penelitian ini adalah data yang digunakan merupakan data balita di Kota Surabaya yang diperoleh dari data Dinas Kesehatan Kota Surabaya.

BAB II
TINJAUAN PUSTAKA

2.1 Klasifikasi

Klasifikasi pada penelitian ini digunakan untuk mengetahui faktor-faktor apa saja yang mengindikasikan terjadinya kasus *stunting* pada balita di Kota Surabaya. Klasifikasi merupakan proses untuk mencari model atau fungsi yang menggambarkan dan mem-bedakan kelas-kelas atau dapat dikatakan sebagai proses pemeriksaan karakteristik objek untuk selanjutnya dapat dike-lompokkan ke dalam salah satu kategori berupa kelas yang sudah didefinisikan sebelumnya (Larose, 2005). Sedangkan menurut Olson (2013), metode-metode klasifikasi ditunjukkan untuk pembelajaran fungsi-fungsi berbeda yang memetakan masing-masing data terpilih ke dalam salah satu dari kelompok kelas yang telah ditetapkan sebelumnya. Proses klasifikasi dapat dibagi menjadi dua fase yaitu fase pembelajaran dan fase klasifikasi. Fase pembelajaran merupakan fase dimana sebagian data yang sudah memiliki label kelas akan diumpamakan untuk membentuk model perkiraan. Pada tahap pembelajaran berfungsi untuk mem-bangun *classifier* dengan menganalisis dari serangkaian data *training*. Selanjutnya pada tahap klasifikasi yang menggunakan data *testing* untuk mengetahui akurasi *classifier* yang terbentuk dari data *training* pada tahap sebelumnya. Dalam klasifikasi kelas biner biasanya terdapat dua kondisi data, yaitu *balance* data dan *imbalance* data. Data yang *imbalance* terjadi saat kondisi data yang tidak seimbang dengan jumlah data suatu kelas melebihi jumlah data kelas lain, kelas data yang banyak disebut kelas mayoritas sedangkan kelas data yang sedikit disebut kelas minoritas. Klasifikasi pada data yang *imbalance* akan cenderung mengklasifikasikan kelas mayoritas dan mengabaikan kelas minoritas. Terdapat beberapa pendekatan untuk mengatasi masa-lah *imbalanced data*, yaitu pendekatan pada level data dengan teknik pengambilan sampel, pendekatan level algoritma, serta metode *ensemble* (Choi, 2010). Teknik pengambilan sampel

yang biasanya digunakan untuk mengatasi masalah data *imbalance* yaitu *oversampling*, *undersampling*, dan kombinasi keduanya. Metode *oversampling* dilakukan untuk menyeimbangkan jumlah distribusi data dengan cara meningkatkan jumlah data kelas minoritas. Sedangkan metode *undersampling* dilakukan dengan mengurangi jumlah data kelas mayoritas agar data menjadi seimbang. Pada penelitian ini dilakukan penanganan data *imbalance* menggunakan metode *oversampling*.

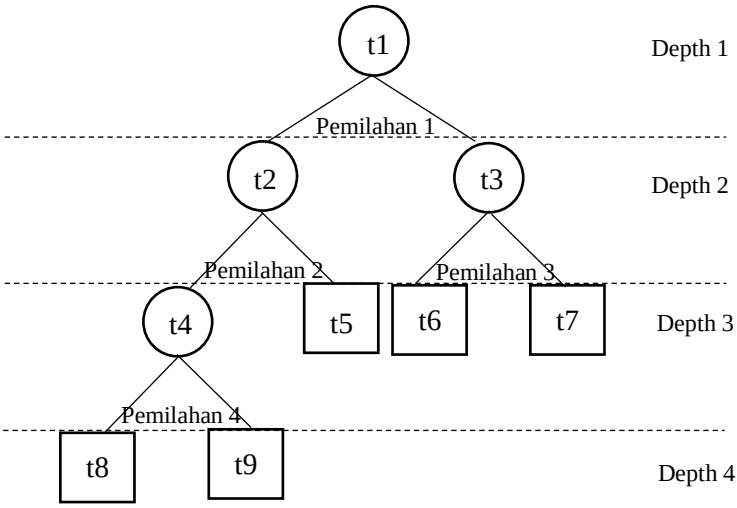
2.2 CART (Classification and Regression Trees)

Metode CART merupakan suatu metodologi statistik untuk analisis klasifikasi, baik untuk variabel dependen berbentuk kategorik maupun kontinu. Metode yang dikembangkan oleh Leo Breiman *et all* ini akan menghasilkan pohon klasifikasi bila variabel dependennya kategorik dan menghasilkan pohon regresi bila variabel dependennya kontinu. Prinsip kerja dari analisis CART disebut sebagai *binary recursive partitioning*. Maksud dari "*binary*" adalah bahwa setiap simpul induk akan dipisah menjadi dua simpul anak, kemudian maksud dari "*recursive*" mengacu pada proses pemecahan yang diulang kembali pada setiap simpul anak sebagai hasil pemecahan terdahulu, sedangkan simpul-simpul anak tersebut sekarang menjadi simpul induk. Proses pemecahan ini akan terus dilakukan sampai tidak ada kesempatan lagi untuk melakukan pemecahan berikutnya. Sedangkan istilah "*partitioning*" mengacu pada data dipisah menjadi bagian-bagian atau partisi-partisi yang lebih kecil (Lewis R. J., 2000). Oleh karena itu setiap simpul induk akan menghasilkan dua simpul anak, dan masing-masing simpul anak akan menjadi simpul induk dan menghasilkan simpul anak, begitu seterusnya. Variabel yang berpengaruh dalam pohon klasifikasi adalah variabel yang menentukan adanya pemilahan (*splitting*) bagi ruang variabel prediktor. Menurut Lewis (2000), terdapat beberapa kelebihan klasifikasi dengan menggunakan metode CART yaitu.

1. Memiliki sifat non parametrik sehingga tidak memerlukan asumsi yang mengikat seperti asumsi distribusi tertentu.

2. Dapat menggunakan kombinasi data kontinu maupun kategorik pada metode ini.
3. Memiliki efisiensi yang baik dalam mengeksplorasi data berdimensi tinggi dengan komputasi.
4. Dapat memperhitungkan interaksi antar variabel prediktor yang berpengaruh. Hal ini berkaitan dengan pengambilan keputusan yang dilakukan secara bertahap dalam himpunan data yang kompleks
5. Tidak hanya memberikan hasil klasifikasi, namun juga memberikan suatu nilai threshold pada setiap pemilah, dan hasil klasifikasi mudah diinterpretasikan.

Pada metode CART, proses analisis digambarkan dalam pohon klasifikasi berbentuk biner sebagai berikut (Breimen, 1993)



Gambar 2. 1 Ilustrasi Pohon Klasifikasi CART

Gambar 2.1 menunjukkan bahwa simpul awal yang mengandung seluruh data dari variabel prediktor yang memiliki hasil evaluasi pemilah tertinggi (*goodness of split* tertinggi) ditandai dengan notasi t1 yang merupakan pemilah terbaik,

kemudian simpul dalam (*internal node*) dinotasikan dengan t2, t3, dan t4 yang mana setiap simpul terdiri dari 1 variabel prediktor, sedangkan simpul akhir (*terminal node*) dinotasikan dengan t5, t6, t7, t8, dan t9 dimana pada simpul terminal tidak ada lagi pemilahan, artinya simpul anak telah memiliki tingkat homogenitas yang tinggi. Pada simpul terminal kemudian dilakukan penandaan kelas dari kelas yang ada pada variabel respon. Setiap simpul berada pada kedalaman (*depth*) tertentu, dimulai dari simpul awal t1 yang berada pada kedalaman 1, t2 dan t3 berada pada kedalaman 2, dan begitu seterusnya hingga dapat simpul t4, t5, t6, t8, dan t9 yang berada pada kedalaman 4.

Terdapat tiga tahap utama dalam proses pembentukan pohon keputusan pada CART. Tahap pertama adalah *tree building process* atau proses pembentukan sebuah pohon klasifikasi, tahap kedua adalah pemberhentian dan pemangkasan pohon klasifikasi, serta tahap ketiga adalah penentuan pohon klasifikasi optimal.

2.2.1 *Tree Building Process*

Tree building process merupakan proses pembentukan atau pembuatan sebuah pohon klasifikasi. Proses ini terdiri dari tiga tahapan yaitu.

a. Proses Pemisahan Simpul

Proses pemisahan simpul dilakukan pada setiap simpul bertujuan mendapatkan pemilah yang mampu menghasilkan simpul dengan tingkat kehomogenan nilai peubah respon paling tinggi. Setiap simpul berada pada kedalaman (*depth*) tertentu, hingga diperoleh hasil dari proses pemilihan yang lebih homogen dibandingkan dengan simpul induknya yang dapat diukur menggunakan nilai *impurity* atau r(t). Aturan pemilahan simpul induk menjadi dua simpul anak *berdasarkan* pada nilai yang berasal dari satu variabel prediktor. Setiap pemilahan hanya bergantung pada satu variabel prediktor saja. Untuk variabel prediktornya yang merupakan variabel kategorik, maka pemilahan berasal dari semua kemungkinan pemilahan berdasarkan terbentuknya dua simpul yang saling lepas (*disjoint*).

Metode pemilihan pemilah menggunakan *impurity measure* atau merupakan pengukuran tingkat keheterogenan suatu kelas dari satu simpul tertentu dalam pohon klasifikasi yang dapat menemukan fungsi pemilah optimal. Fungsi *impurity measure* yang digunakan adalah fungsi indeks gini, hal ini karena kelebihan yang dimiliki fungsi indeks gini yaitu proses perhitungannya sederhana dan relatif cepat, serta mudah untuk diterapkan dalam berbagai kasus. Berikut merupakan fungsi indeks gini yang sering digunakan.

$$r(t) = \sum_{i \neq j} p(j \mid t)\, p(i \mid t), \qquad (2.1)$$

keterangan :

$r(t)$: indeks gini (fungsi heterogenan) pada simpul *t*

$p(i \mid t)$: proporsi kelas *i* pada simpul *t*

$p(j \mid t)$: proporsi kelas *j* pada simpul *t*

Kemudian dilanjutkan dengan menentukan pemilah terbaik dari setiap variabel prediktor, dimana pemilah terbaik merupakan pemilah yang memaksimalkan ukuran homogenitas setiap simpul anak terhadap simpul induknya dan juga untuk dapat memaksimalkan ukuran pemilahan antara dua simpul anak tersebut. Setiap pemilahan akan dilakukan pada setiap simpul sampai diperoleh simpul akhir. Kemudian, menentukan kriteria *goodness of split* yang merupakan suatu evaluasi pemilahan yang dilakukan oleh pemilah s pada suatu simpul t sebagai berikut.

$$\varnothing(s,t) = \Delta i(s,t) = r(t) - p_L(r(t_L)) - p_R(r(t_R)), \qquad (2.2)$$

keterangan :

$\varnothing(s,t)$: nilai *goodness of split*

$r(t)$: fungsi heterogenitas pasa simpul *t*

p_L : proporsi pengamatan simpul kiri

p_R : proporsi pengamatan simpul kanan

$r(t_L)$: fungsi heterogenitas simpul kiri

$r(t_R)$: fungsi hetergenitas simpul kanan

Berikut merupakan tabel ilustrasi yang digunakan untuk menghitung proporsi pengamatan simpul.

Tabel 2. 1 Ilustrasi Tabel untuk Perhitungan Proporsi Pengamatan Simpul

Prediktor	Respon		Total
	0	1	
0	n_{00}	n_{01}	N_0
1	n_{10}	n_{11}	N_1
			N_{Total}

p_L merupakan nilai proporsi pengamatan simpul kiri yang diperoleh dari $\frac{N_0}{N_{Total}}$, N_0 merupakan jumlah banyak data pada variabel prediktor kategori 0 yang berada di simpul kiri (jika variabel prediktor kategori 0 berada di simpul kiri), kemudian p_R merupakan nilai proporsi pengamatan simpul kanan yang diperoleh dari $\frac{N_1}{N_{Total}}$, N_1 merupakan jumlah banyak data pada variabel prediktor kategori 1 yang berada di simpul kanan. Pemilah yang menghasilkan nilai *goodness of split* tertinggi merupakan pemilah terbaik, karena mampu menurunkan nilai heterogenitas atau nilai impuritas yang tinggi. Kemudian dilakukan pengulangan langkah tersebut untuk menentukan variabel yang digunakan sebagai pemilah *node*. Setiap variabel akan menghasilkan nilai atau skor untuk menunjukkan seberapa besar variabel tersebut memberikan kontribusi dalam proses pembentukan pohon. Nilai skor diperoleh dengan menjumlahkan nilai *goodness of split* dari variabel yang berperan sebagai pengganti untuk setiap simpul. Berikut merupakan rumus untuk menentukan besar Skor tiap variabel.

$$\text{Skor} = \sum_{g=1}^{G} \varnothing(s,t_g)\,, \tag{2.3}$$

$\varnothing(s,t_g)$ merupakan nilai *goodness of split* pada setiap simpul dimana t_g merupakan indeks dari simpul dan G merupakan banyak simpul.

b. Penentuan Simpul Terminal

Suatu simpul t ditentukan menjadi simpul terminal atau tidak akan dipilah kembali bila pada simpul t tidak terdapat penurunan keheterogenan secara berarti atau adanya batasan minimum n seperti hanya terdapat satu pengamatan pada tiap simpul anak. Jumlah kasus minimum dalam suatu terminal akhir umumnya adalah 5, dan apabila hal itu terpenuhi maka pengembangan pohon dihentikan (Breiman L., 1993).

c. Penandaan Label Kelas

Penandaan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak. Proses penandaan label kelas merupakan proses identifikasi simpul terminal pada suatu kelas tertentu. Label kelas untuk simpul terminal *t* adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian paling kecil sebesar $r(t) = 1 - \max_j p(j \mid t)$.

$$p(j_0 \mid t) = \max_j p(j \mid t) = \max_j \frac{N_j(t)}{N(t)}, \qquad (2.4)$$

keterangan :

$p(j_0 \mid t)$: proporsi kelas *j* pada simpul terminal t

$N_j(t)$: jumlah pengamatan kelas *j* pada simpul *t*

$N(t)$: jumlah pengamatan pada simpul *t*

2.2.2 Pemangkasan Pohon Klasifikasi

Proses pemangkasan pohon klasifikasi perlu dilakukan adalah untuk menghindari terjadinya *overfitting* dimana semakin banyak pemilahan yang dilakukan mengakibatkan makin kecilnya tingkat kesalahan prediksi atau dengan kata lain nilai prediksi melebihi nilai yang sebenarnya. Pemangkasan didasarkan pada suatu penilaian ukuran sebuah pohon tanpa mengorbankan kebaikan ketepatan melalui pengurangan simpul pohon sehingga dicapai ukuran pohon yang layak. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak tersebut adalah *cost complexity minimum* (Breiman, 1993).

$$R_\alpha(T) = R(T) + \alpha\,|\tilde{T}|, \qquad (2.5)$$

keterangan :

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

$R(T)$: ukuran kesalahan klasifikasi pohon T

α : parameter *cost complexity* bagi penambah satu simpul terminal pada pohon T

$\tilde{T}$: banyak simpul terminal pada pohon T

Cost complexity pruning digunakan untuk menentukan pohon bagian T (a) yang dapat meminimumkan (T) pada pohon bagian atau setiap nilai a. Nilai parameter kompleksitas (a) akan secara perlahan meningkat selama proses pemangkasan. Selanjutnya, pencarian pohon bagian T $(a) < T_{\text{maks}}$ yang dapat meminimumkan $R_\alpha(T)$. Pemangkasan dilakukan berdasarkan *weakest-link cutting* dimulai dari pohon T_1 dengan nilai kritis dari α menggunakan persamaan sebagai berikut.

$$g_1(t) = \frac{R(t) - R(T_t)}{\left|\tilde{T}_t\right| - 1}, t \notin \tilde{T}_1 \qquad (2.6)$$

kemudian *weakest-link* $\bar{t}_1$ pada T_1 adalah

$$g_1(\bar{t}_1) = \min_{t \in T_1} g_1(t) \qquad (2.7)$$

dimana

$$\alpha_2 = g_1(\bar{t}_1) \qquad (2.8)$$

2.2.3 Penentuan Pohon Klasifikasi Optimum

Besarnya ukuran pohon akan menyebabkan nilai kompleksitas yang tinggi karena struktur data yang digambarkan cenderung komplek, sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti cukup kecil. Pemilihan pohon dapat langsung menggunakan pohon yang meminimumkan (T) atau dengan membentuk estimasi tak bias dari *missclassification cost* $R^*(T_t)$. Metode yang sering digunakan adalah *test sample estimate*, dimana penduga

sampel uji L dibagi menjadi dua himpunan data, yaitu L_1 (*learning set)* dan L_2 (*testing set*). Data yang masuk kedalam kelompok L_1 digunakan untuk pembentukan pohon klasifikasi. Data yang masuk kedalam kelompok L_2 digunakan untuk menduga $R^{ts}(T_t)$ dimana $N^{(2)}$ adalah jumlah amatan dalam L_2, dan untuk setiap pohon $T_1, T_2, ...,$ $N_{ij}^{(2)}$ adalah jumlah *instance* kelas j dalam data *testing* yang diprediksi menjadi kelas i. Pohon optimal adalah T^* yang memenuhi kriteria $R^{ts}(T^*) = \min R^{ts}(T_t)$. *Expexted missclassification cost* untuk *classifier d* atau total kesalahan klasifikasi *test sample* oleh T dinyatakan oleh persamaan (2.9).

$$R^{ts}(T_t) = \frac{1}{N^{(2)}} \sum_{i,j} c(i \mid j) N_{ij}^{(2)} \qquad (2.9)$$

2.3 *Value Difference Metric* (VDM)

Value Different Metric diperkenalkan oleh Stanfill dan Waltz pada tahun 1986 untuk memberikan fungsi perhitungan jarak yang tepat untuk data dengan atribut kategorikal. VDM ini mendefinisikan jarak antar kategori dari suatu variabel. Jarak antar kategori dalam suatu variabel dijelaskan oleh persamaan (2.10).

$$\delta(x_j, y_j) = \sum_{i=1}^{n} \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2}, \right|^k \qquad (2.10)$$

dimana :

$\delta(x_j, y_j)$: jarak antara kategori x_j dan y_j

C_1 : banyak x_j terjadi

C_2 : banyak y_j terjadi

C_{1i} : banyak x_j yang masuk kelas i

C_{2i} : banyak y_j yang masuk kelas i

k : konstan (digunakan 1)

Nilai k yang biasa digunakan adalah 1 atau 2, namun digunakan 1 agar memperoleh *distance* terdekat dikarenakan k=1 yang

meminimumkan jarak antara ketegori x_j dan y_j. Selanjutnya Jarak antar nilai dihitung menggunakan persamaan (2.11).

$$\Delta(P,Q) = w_p w_q \sum_{i=1}^{p} \delta(x_j, y_j)^r \quad , \qquad (2.11)$$

dimana :

$\Delta(P,Q)$	: jarak antara suatu nilai *P* dan *Q*
w_p, w_q	: bobot (dapat diabaikan)
p	: banyak variabel bebas
x_j	: data X pada variabel ke-*i*
y_j	: data Y pada variabel ke-*i*
r	: 1 (Manhattan) dan 2 (Euclidean)

Apabila nilai *r* yang digunakan adalah 2 yang merupakan jarak euclidean, maka jarak antara suatu nilai *P* dan *Q* adalah akar dari jarak antara kategori x_j dan y_j yang telah dikuadratkan.

Jarak pada ADASYN-N dihitung menggunakan versi modifikasi dari *Value Difference Metric* (VDM) seperti pada SMOTE-N (Rahayu, 2017). VDM melihat pada nilai fitur yang *overlap* terhadap semua vektor fitur. Matriks mendefinisikan jarak antara nilai fitur yang sesuai untuk vektor fitur yang dibuat.

2.4 *Adaptive Synthetic Nominal* (ADASYN-N)

ADASYN merupakan metode untuk pendekatan sampling pada pembelajaran dengan dataset yang tidak seimbang. Konsep dari ADASYN adalah memberikan bobot pada data dalam kelas minoritas berdasarkan pada tingkat kesulitan belajar, sehingga data sintetis dihasilkan dari kelas minoritas yang susah untuk belajar dibandingkan dengan data minoritas yang lebih mudah untuk belajar. ADASYN meningkatkan pembelajaran dengan dua cara. Pertama, mengurangi bias yang diakibatkan oleh ketidakseimbangan kelas dan yang kedua secara adaptif menggeser batas keputusan klasifikasi terhadap kesulitan data.

Kemudian pada tahun 2017, Kurniawati mengembangkan ADASYN menjadi ADASYN-N melalui pendekatan data dengan tipe nominal. Jarak pada ADASYN-N dihitung menggunakan versi modifikasi dari *Value Difference Metric* (VDM) seperti pada SMOTE-N (Rahayu, 2017).

Pada ADASYN-N, *training dataset* dengan *m* sampel $\{x_i, y_i\}, i = 1, \dots, m$ dimana x_i adalah *instance* dalam *n* dimensional *feature space* $\mathbf{X}$ atau bisa dikatakan *instance* dari tiap variabel prediktor pada *training dataset*, dan $y_i \in Y = \{1, \dots, C\}$ adalah label identitas kelas dengan jumlah *instance* terbanyak. Kemudian data dibagi menjadi m_s dan m_l, dimana merupakan jumlah *instance* kelas minoritas dan mayoritas. Oleh karena itu $m_s \leq m_l$ dan $\sum m_s + m_l = \mathrm{m}$. Selanjutnya dilakukan perhitungan untuk *degree of class imbalance* seperti pada persamaan (2.12).

$$d = \frac{m_s}{m_l} \tag{2.12}$$

Selanjutnya untuk perhitungan jumlah *instance* data sintetis yang perlu digeneralisasi untuk kelas minoritas digunakan persamaan (2.13).

$$G = (m_l - m_s) \times \beta \tag{2.13}$$

dimana $\beta \epsilon [0,1]$ merupakan parameter yang digunakan untuk menetapkan *level balance* yang diinginkan setelah generalisasi data sintetis, nilai β=1 menunjukkan data yang sepenuhnya seimbang setelah proses generalisasi oleh algoritma ADASYN dimana nilai β=1 memiliki *classification error performance* yang paling rendah (He, 2008). Jadi nilai β telah di*define* dari awal sebesar 1 untuk memperoleh level *balance* yang diinginkan. Pada setiap *instance* $x_i \in$ *minority class*, kemudian ditentukan k-*nearest neighbors* pada *n dimensional space*, dan kalkulasi rasio r_i sebagai berikut.

$$r_i = \frac{\Delta_i}{k}, i = 1, \dots, m_s \tag{2.14}$$

Dimana r_i merupakan dominasi kelas mayoritas pada masing-masing tetangga dan Δ_i merupakan jumlah *instance* dalam *nearest neighbor* yang termasuk kelas mayoritas, dan k merupakan nilai k dari *nearest neighbor*. Selanjutnya dilakukan normaliasasi r_i dengan persamaan (2.15).

$$\hat{r}_i = \frac{r_i}{\sum_{i=1}^{m_s} r_i} \tag{2.15}$$

dimana

$$\sum_{i=1}^{m_s} \hat{r}_i = 1 \tag{2.16}$$

Kemudian dilakukan perhitungan jumlah *instance* sintetis yang perlu dihasilkan bagi setiap *instance* minoritas menggunakan persamaan (2.17).

$$g_i = \hat{r}_i \times G \tag{2.17}$$

Setelah memperoleh jumlah data yang harus dibangkitkan bagi setiap *instance* pada kelas minoritas, dilakukan replikasi pada masing-masing *instance* sebanyak g_i.

2.5 *Adaptive Synthetic*-kNN (ADASYN-kNN)

ADASYN-kNN merupakan pengembangan dari ADASYN-N dimana dilakukan pengembangan pada prosedur untuk menghasilkan *instance* data sintetis sebanyak g_i. Pada metode ini dilakukan dilakukan perhitungan untuk *degree of class imbalance*. Selanjutnya dilakukan perhitungan jumlah *instance* data sintetis yang perlu digeneralisasi untuk kelas minoritas. Pada setiap *instance* $x_i \in$ *minority class*, kemudian ditentukan k-*nearest neighbors* pada *n dimensional space* serta dilakukan kalkulasi rasio r_i kemudian dilakukan normalisasi r_i, setelah dilakukan normalisasi r_i kemudian dicari *nearest neighbor* dari *instance* minoritas, kemudian dilakukan *majority voting* dari atribut *nearest neighbor instance* minoritas tersebut. Kemudian *instance* baru dihasilkan berdasarkan pada *majority voting*, dan hasil *instance* baru tersebut diduplikasi sebanyak g_i.

2.6 Ketepatan Klasifikasi

Setelah dilakukan pembagian dataset yaitu membagi dataset menjadi dua bagian, bagian pertama digunakan sebagai data *training* dan bagian kedua digunakan sebagai data *testing* yang dipakai untuk memvalidasi model, selanjutnya dilakukan pengukuran ketepatan klasifikasi untuk mengetahui seberapa besar kemampuan suatu metode dalam mengklasifikasikan data ke dalam kelas yang tepat. Terdapat beberapa cara yang sering digunakan dalam mengukur ketepatan klasifikasi antara lain akurasi, sensitifitas, dan spesifisitas. Selain itu juga ada *confusion matrix* atau disebut juga matriks klasifikasi yang merupakan salah satu alat visual yang biasanya digunakan dalam *supervised learning*. Matriks klasifikasi berisi jumlah kasus-kasus yang diklasifikasikan dengan benar dan kasus-kasus yang salah diklasifikasikan. Pengukuran ketepatan klasifikasi dirumuskan berdasarkan tabel *confusion matrix* berikut.

Tabel 2. 2 Ketepatan Klasifikasi

Kelas Aktual	**Kelas Prediksi**	
	Positif	Negatif
Positif	*True Positive* (TP)	*False Negative* (FN)
Negatif	*False Positive* (FP)	*True Negative* (TN)

Berikut merupakan rumus untuk mengukur ketepatan klasifikasi.

$$\text{sensitifitas} = \frac{TP}{TP + FN} \tag{2.18}$$

$$\text{spesifisitas} = \frac{TN}{TN + FP} \tag{2.19}$$

Setelah memperoleh nilai akurasi, sensitifitas, dan spesifisitas, selanjutnya dilakukan perhitungan *Area Under Curve* (AUC). AUC merupakan indikator performasi kurva *Receiver Operating Characteristic* (ROC) yang dapat meringkas kinerja sebuah *classifier* menjadi satu nilai. AUC merupakan evaluasi prediktor yang lebih komprehensif dalam konteks ketidakseimbangan atau data *imbalance* (Wang & Yao, 2013)

$$AUC=\frac{sensitifitas+spesifitas}{2} \quad (2.20)$$

Untuk interval ketepatan klasifikasi yang baik dengan perhitungan AUC terdapat pada Tabel 2.3 berikut (Bekkar, Djemaa, & Alitouche, 2013).

Tabel 2. 3 Kriteria Nilai AUC

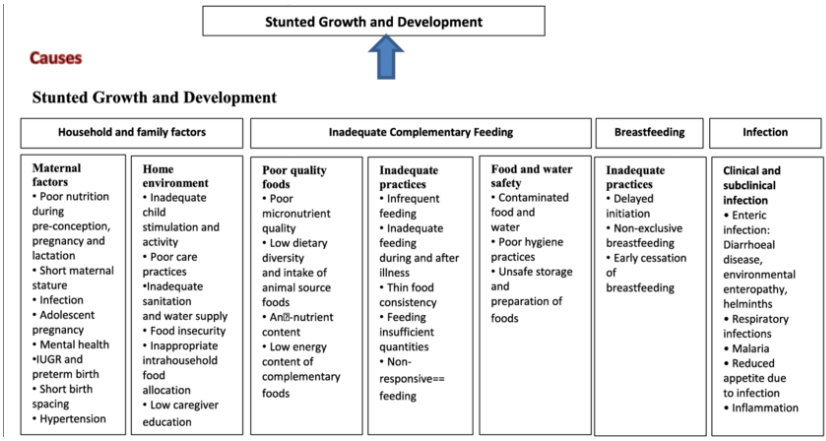
Nilai AUC	Keterangan
0,5-0,6	Kurang baik
0,6-0,7	Cukup
0,7-0,8	Baik
0,8-0,9	Sangat Baik
0,9-1,0	Sempurna

2.7 Balita *Stunting*

Stunting adalah kondisi dimana balita memiliki panjang atau tinggi badan yang kurang jika dibandingkan dengan umur. Kondisi ini diukur dengan panjang atau tinggi badan yang lebih dari minus dua standar deviasi median standar pertumbuhan anak dari WHO. Sedangkan menurut UNICEF, *stunting* merupakan persentase anak-anak usia 0-59 bulan berdasarkan *z score* dari perhitungan tinggi badan atau panjang badan dibandingkan umur. *Stunting* dikelompokkan menjadi pendek dan sangat pendek. Pendek atau *stunted* adalah status gizi balita TB/U yang memiliki *z score* antara -3 SD hingga –2SD, sedangkan sangat pendek atau *severely stunted* adalah status gizi balita TB/U yang nilainya kurang dari -3SD (Kementerian Kesehatan, 2010). *Z-score* adalah nilai simpangan TB dari nilai TB normal menurut baku pertumbuhan WHO. Apabila balita memiliki status gizi sangat pendek atau pendek pada indikator tinggi badan dibandingan dengan umur (TB/U), maka balita tersebut dikategorikan *stunting*. Balita *stunting* termasuk masalah gizi kronik yang disebabkan oleh banyak faktor seperti kondisi sosial ekonomi, gizi ibu saat hamil, kesakitan pada bayi, dan kurangnya asupan gizi pada bayi (Kemenkes RI, 2018). *Stunting* terjadi mulai dari dalam

kandungan dan baru terlihat saat anak berusia dua tahun. *Stunting* menimbulkan gangguan perkembangan fisik anak yang *irreversible*, menyebabkan penurunan kemampuan kognitif dan motorik serta penurunan performa kerja. Anak *stunting* memiliki rata-rata skor *Intelligence Quotient* (IQ) sebelas poin lebih rendah dibandingkan rata-rata skor IQ pada anak normal. Gangguan tumbuh kembang pada anak akibat kekurangan gizi kronis ini bila tidak mendapatkan intervensi sejak dini akan berlanjut hingga dewasa (Trihono, 2015). *Stunting* pada anak juga berhubungan dengan peningkatan kerentanan anak terhadap penyakit, baik penyakit menular maupun Penyakit Tidak Menular (PTM) serta peningkatan risiko *overweight* dan obesitas. Keadaan *overweight* dan obesitas jangka panjang dapat meningkatkan risiko penyakit *degenerative* (Setiawan, dkk, 2018).

World Health Organization (WHO) mempublikasikan *conceptual framework* yang berisi faktor-faktor yang menyebabkan terhambatnya pertumbuhan dan perkembangan balita (penyebab balita *stunting*). Terdapat 4 faktor utama yaitu faktor keluarga dan rumah tangga, faktor makanan pelengkap yang tidak memadai, faktor menyusui, serta faktor infeksi. Menurut WHO, diantara faktor-faktor penyebab langsung *stunting*, pemberian makanan pelengkap disoroti bersama dengan pentingnya menyusui. Faktor ibu memiliki kemungkinan akan berdampak langsung terhadap pertumbuhan dan perkembangan keturunan pada 1000 hari pertama. Bagian lingkungan rumah juga merupakan hal penting yang perlu diperhatikan, rumah yang ideal menyediakan lingkungan yang mendukung, bersih, aman dan sesuai dengan keadaan serta kebutuhan ibu dan anak. Berikut merupakan *conceptual framework* dari World Health Organization (WHO).



Gambar 2. 2 WHO *Conceptual Framework of Stunting*

Pada penelitian ini dilakukan analisis klasifikasi terhadap faktor-faktor yang diduga berpengaruh terhadap kasus *Stunting* pada balita di Kota Surabaya. Variabel yang digunakan mengacu pada faktor keluarga dan rumah tangga, faktor kelengkapan makanan yang tidak memadai, faktor menyusui, serta faktor infeksi. Beberapa variabel yang digunakan pada faktor keluarga dan rumah tangga antara lain.

– Usia Kehamilan

Kehamilan cukup bulan merupakan kehamilan usia 37-42 minggu (Prawirohardjo, 2006).

– Status Gizi Balita

Status gizi balita dinilai menurut tiga indeks, yaitu Berat Badan Menurut Umur (BB/U), Tinggi Badan Menurut Umur (TB/U), dan Berat Badan Menurut Tinggi Badan (BB/TB). BB/U adalah berat badan anak yang dicapai pada umur tertentu. TB/U adalah tinggi badan anak yang dicapai pada umur tertentu. Kemudian BB/TB adalah berat badan anak dibandingkan dengan tinggi badan yang dicapai (Kementrian Kesehatan, 2010). Untuk tabel pengertian kategori status gizi balita tersebut ditunjukkan pada Tabel 2.4.

Tabel 2. 4 Pengertian Kategori Status Gizi Balita

Indikator	Status Gizi	Z-Score
BB/U	Gizi Buruk	<-3,0 SD
	Gizi Kurang	-3,0 SD s/d -2,0 SD
	Gizi Baik	-2,0 SD s/d 2,0 SD
	Gizi Lebih	>2,0 SD
TB/U	Sangat Pendek	<-3,0 SD
	Pendek	-3,0 SD s/d -2,0 SD
	Normal	≥-2,0 SD
BB/TB	Sangat Kurus	<-3,0 SD
	Kurus	-3,0 SD s/d -2,0 SD
	Normal	-2,0 SD s/d 2,0 SD
	Gemuk	>2,0 SD

Z-Score merupakan nilai simpangan BB atau TB dari nilai BB atau TB normal menurut baku pertumbuhan WHO. Pada indikator TB/U, apabila balita memiliki status gizi sangat pendek atau pendek, maka balita tersebut dikategorikan *stunting*.

– Penyakit Kronis Balita

Tingkat pendidikan ibu yang tinggi, jarak kelahiran lebih dari 60 bulan, berat lahir normal, dan riwayat penyakit kronis merupakan faktor protektif malnutrisi pada balita (Kuntari dkk, 2013).

– Total Pendapatan Keluarga

Status ekonomi keluarga yang rendah akan mempengaruhi kualitas maupun kuantitas bahan makanan yang dikonsumsi oleh keluarga. Makanan yang didapat biasanya akan kurang bervariasi dan sedikit jumlahnya terutama pada bahan pangan yang berfungsi untuk pertumbuhan anak seperti sumber protein, vitamin dan mineral, sehingga meningkatkan risiko kurang gizi. Keterbatasan ini akan meningkatkan risiko terjadinya

Stunting pada anak balita (Astari, 2006). Status ekonomi diukur dari pendapatan keluarga karena ketidakmampuan dalam keuangan menyebabkan kurangnya kemampuan keluarga untuk memenuhi asupan gizi keluarga sesuai dengan kebutuhan yang seharusnya (Sudarsih & Wijayanti, 2013).

– Jumlah Anak dalam Keluarga

Menurut BKKBN jumlah anggota rumah tangga yang disarankan adalah empat orang, dimana terdiri dari ayah, ibu, dan dua orang anak. Jumlah anggota keluarga yang banyak kurang baik bagi anak-anak. Keluarga yang jumlah anggotanya lebih banyak, disertai dengan pendapatan keluarga yang rendah, maka anggota keluarga tersebut terutama anak-anak berpeluang untuk tidak mendapat asupan yang lebih baik (Lestari, dkk, 2018).

– Anggota Keluarga Ada yang Merokok dalam Rumah

Perilaku merokok pada orangtua diperkirakan memiliki pengaruh terhadap terjadinya kasus *stunting* pada anak. Asap rokok orang tua perokok dapat memberi efek langsung pada tumbuh kembang anak. Selain itu, apabila melihat sisi biaya belanja rokok, besar kemungkinan orang tua mengurangi “jatah” biaya belanja makanan bergizi, biaya kesehatan, pendidikan dan seterusnya (Kemenkes RI, 2018).

– Sarana Air Bersih

Menurut definisi operasional rumah sehat Dinas Kesehatan Kota Surabaya, suatu sarana air bersih dikatakan layak apabila merupakan milik sendiri dan memenuhi syarat kesehatan, dan dikatakan tidak layak sebagai air bersih apabila bukan milik sendiri dan memenuhi syarat kesehatan/milik sendiri dan tidak memenuhi syarat kesehatan.

Sedangkan beberapa variabel yang digunakan dan mengacu pada faktor kelengkapan makanan yang tidak memadai adalah.

– Pemberian Makanan Tambahan

Makanan Tambahan Balita adalah suplementasi gizi berupa makanan tambahan dalam bentuk biskuit dengan formulasi khusus dan difortifikasi dengan vitamin dan mineral

yang diberikan kepada bayi dan anak balita usia 6-59 bulan dengan kategori kurus untuk mencukupi kebutuhan gizi. Pemberian makanan tambahan ini merupakan salah satu strategi suplementasi dalam mengatasi masalah gizi. Pemberian makanan tambahan yang berfokus baik pada zat gizi makro maupun zat gizi mikro bagi sangat diperlukan dalam mencegah terjadinya kasus *stuting* pada balita (Kemenkes RI, 2018).

– Pemberian CFC/TFC

Therapeutic Feeding Center (TFC) atau Pusat Pemulihan Gizi berfungsi sebagai tempat perawatan dan pengobatan anak gizi buruk secara intensif di suatu tempat/ruangan khusus dan ibu/ keluarga ikut aktif terlibat dalam perawatan anak gizi buruk. Umumnya perawatan terhadap anak balita dengan gizi buruk di tempat ini antara 2-3 bulan. Ibu balita menginap sambil menambah ilmu bagaimana merawat anak balitanya, cara memasak dengan pemberian makanan mengandung tinggi kalori dan protein dengan aneka bahan makanan setempat sehingga kekurangan BB terpenuhi dan dapat meningkatkan tinggi badan. *Community Feeding Center* (CTC) adalah pos pemulihan gizi berbasis masyarakat yang merupakan upaya masyarakat untuk memantau atau merawat anak balita. Rujukan balik dari Puskesmas agar status gizinya tetap normal. Pos Gizi ini merupakan salah satu kegiatan pengembangan Posyandu (Posyandu Plus) (Andewi, 2011).

– Pemberian Kapsul Vitamin A

Pemberian Kapsul Vitamin A merupakan pembagian suplementasi vitamin A Kapsul biru (dosis 100.000 IU) untuk bayi umur 6-11 bulan dan kapsul merah (dosis 200.000 IU) untuk anak umur 12-59 bulan yang umumnya dilakukan pada Bulan Februari dan Agustus (Kemenkes RI, 2019).

Selanjutnya untuk variabel yang mengacu pada faktor menyusui adalah pemberian ASI eksklusif dan inisiasi menyusui dini.

– Pemberian ASI Eksklusif

ASI eksklusif adalah ASI yang diberikan kepada bayi sejak dilahirkan selama enam bulan tanpa menambahkan dan/atau mengganti dengan makanan atau minuman lain (Satriani S, 2018).

– Inisiasi Menyusui Dini (IMD)

Inisiasi Menyusui Dini adalah permulaan kegiatan menyusu dalam satu jam pertama setelah bayi lahir. Inisiasi dini juga merupakan cara bayi menyusu satu jam pertama setelah lahir dengan usaha sendiri (Maryuni, 2012). Saat melakukan IMD, bayi memperoleh kolostrom yang mengandung antibodi dan merupakan imunisasi pertama, selain itu kolostrom juga mengandung faktor pertumbuhan yang membantu usus bayi berfungsi secara efektif, sehingga mikroorganisme dan penyebab alergi lain akan lebih sulit masuk ke dalam tubuh bayi (Roesli, 2008).

Faktor selanjutnya yaitu faktor kelengkapan imunisasi dasar. Pada faktor ini terdapat variabel kelengkapan imunisasi dimana yang dimaksud dengan imunisasi dasar pada bayi adalah upaya pencegahan terjangkitnya penyakit tertentu yaitu penyakit yang dapat dicegah dengan imunisasi (PD3I), antara lain Tuberkulosis, Difteri, Pertusis, Tetanus, Hepatitis B, Polio, dan Campak (Kemenkes RI, 2014).

BAB III
METODOLOGI PENELITIAN

3.1 Sumber Data

Sumber data yang akan digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari Dinas Kesehatan Kota Surabaya. Data merupakan hasil dari survei pendampingan balita di Kota Surabaya pada Agustus 2019 hingga Februari 2020 kerjasama Dinas Kesehatan Kota Surabaya dengan UPT Laboratorium Gizi.

3.2 Variabel Penelitian dan Definisi Operasional

Variabel yang digunakan dalam penelitian ini terdiri atas 1 variabel dependen dan 26 variabel independen yang merupakan variabel pada data survei pendampingan balita di Kota Surabaya oleh Dinas Kesehatan Kota Surabaya. Keterangan lebih lanjut mengenai variabel penelitian yang digunakan ditunjukkan oleh Tabel 3.1.

Tabel 3. 1 Variabel Penelitian

No	Nama Variabel	Keterangan
Variabel Terikat		
1	Status Balita (Y)	Kategori 0. *Stunting* 1. Normal
Variabel Bebas		
Faktor Keluarga dan Rumah Tangga		
	Faktor Maternal	
2	Riwayat Penyakit Ibu Saat Hamil (X1)	Kategori 0. Tidak Ada 1. Hipertensi 2. Diabetes Mellitus 3. Pre Eklamsia 4. Ketuban Pecah Dini 5. Lainnya

Tabel 3. 1 Variabel Penelitian (Lanjutan)

No	Nama Variabel	Keterangan
3	Usia Kehamilan (X2)	Kategori 0. < 37 Minggu 1. ≥37 Minggu
4	Jenis Kelamin Balita (X3)	Kategori 0. Perempuan 1. Laki-laki
5	Kategori BB/TB (X4)	Kategori 0. Sangat Kurus 1. Kurus 2. Normal
6	Kategori BB/U (X5)	Kategori 0. Gizi Buruk 1. Gizi Kurang 2. Gizi Baik 3. Gizi Lebih
7	Kategori TB/U (X6)	Kategori 0. Sangat Pendek 1. Pendek 2. Normal
8	Penyakit Kronis Balita (X7)	Kategori 0. Tidak Ada 1. Kelainan Jantung Bawaan 2. Bibir Sumbing 3. Down Syndrome 4. Cerebral Palsy 5. Kaki Bengkok 6. Pneumonia 7. HIV/AIDS 8. TBC 9. Lainnya

Tabel 3. 1 Variabel Penelitian (Lanjutan)

No	Nama Variabel	Keterangan
	Faktor Lingkungan dan Rumah Tangga	
9	Tingkat Pendidikan Ibu (X8)	Kategori 0. Tidak Ada Jawaban (Meninggal dunia/tidak tahu,dll) 1. Tidak Lulus SD 2. Lulus SD 3. Lulus SMP 4. Lulus SMA 5. Lulus PT (D1/D2/D3/S1/S2/S3)
10	Status Bekerja Ibu (X9)	Kategori 0. Tidak Bekerja 1. Bekerja
11	Total Pendapatan Keluarga (X10)	Kategori 0. <UMK 1. ≥UMK
12	Jumlah Anak dalam Keluarga (X11)	Kategori 0. >2 1. ≤ 2
13	Status Balita (X12)	Kategori 0. Tidak Pernah Didampingi 1. Pernah Didampingi
14	Anggota Keluarga Ada yang Merokok dalam Rumah (X13)	Kategori 0. Tidak 1. Ya
15	Keteraturan ke Posyandu (X14)	Kategori 0. Tidak Teratur 1. Teratur
16	Makan Buah Sayur Setiap Hari (X15)	Kategori 0. Tidak 1. Ya

Tabel 3. 1 Variabel Penelitian (Lanjutan)

No	Nama Variabel	Keterangan
17	Jamban (X16)	Kategori 0. Jamban Sehat Permanen 1. Jamban Sehat Semi Permanen 2. Jamban Umum 3. BAB Sembarangan
18	Tempat Sampah (X17)	Kategori 0. Tidak Ada 1. Tidak Memenuhi Syarat 2. Memenuhi Syarat
19	Sarana Air Bersih (X18)	Kategori 0. Tidak Ada 1. Tidak Memenuhi Syarat 2. Memenuhi Syarat
Faktor	Kelengkapan Makanan yang Tidak Memadai	
	Makanan Pelengkap	
20	Pemberian Makanan Tambahan (X19)	Kategori 0. Tidak 1. Ya
21	Pemberian Taburia (X20)	Kategori 0. Tidak 1. Ya
22	Pemberian CFC/TFC (X21)	Kategori 0. Tidak 1. Ya
23	Pemberian kapsul Vit. A (X22)	Kategori 0. Tidak Teratur 1. Teratur

Tabel 3. 1 Variabel Penelitian (Lanjutan)

No	Nama Variabel	Keterangan
Keamanan Air dan Makanan		
24	Sarana Air Minum (X23)	Kategori 0. Kurang Baik 1. Baik
Faktor Menyusui		
25	Pemberian ASI Eksklusif (X24)	Kategori 0. Tidak (diberikan cairan/makanan sebelum usia 6 bulan) 1. Ya (hanya diberikan ASI hingga usia 6 bulan)
26	Inisiasi Menyusui Dini (X25)	Kategori 0. Tidak 1. Ya
Faktor Infeksi		
27	Status Kelengkapan Imunisasi (X26)	Kategori 0. Tidak lengkap 1. Lengkap

Data yang digunakan pada penelitian ini merupakan data jadi yang diperoleh dari Dinas Kesehatan Kota Surabaya sehingga penentuan status *stunting* atau normal, status pada variabel TB/U, BB/U, BB/TB, dan variabel lain telah dilakukan oleh Dinas Kesehatan Kota Surabaya.

3.3 Struktur Data

Berikut merupakan struktur data dari variabel-variabel yang digunakan dalam penelitian.

Tabel 3. 2 Struktur Data Penelitian

No	X_1	X_2	X_3	…	X_k	Y
1	X_{11}	X_{21}	X_{31}	…	X_{k1}	Y_{11}
2	X_{12}	X_{22}	X_{32}	…	X_{k2}	Y_{12}

Tabel 3.2 Struktur Data Penelitian (Lanjutan)

No	X_1	X_2	X_3	…	X_k	Y
⋮	⋮	⋮	⋮		⋮	⋮
n	X_{1n}	X_{2n}	X_{3n}	…	X_{kn}	Y_{1n}

dimana X_k merupakan variabel ke-k dan Y_n merupakan kelas pada data ke-n.

3.4 Langkah Analisis

Langkah analisis digunakan untuk menggambarkan langkah-langkah penelitian yang akan dilakukan secara urut. Langkah analisis yang digunakan adalah sebagai berikut.

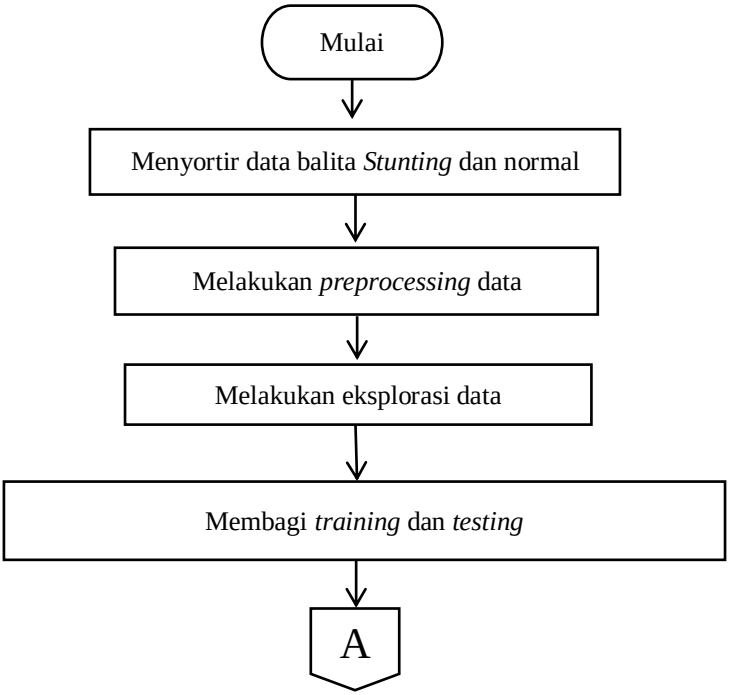
1. Melakukan pemilahan pada data balita *stunting* di Kota Surabaya. Data balita dikelompokkan menjadi dua kelas yaitu kelas normal dan kelas *stunting*.
2. Melakukan *preprocessing* data.
3. Melakukan eksplorasi data variabel dependen dan independen.
4. Membagi data *training* dan *testing*
5. Melakukan penanganan terhadap data *imbalance* menggunakan metode ADASYN-N.
   a. Menghitung jumlah data sintetis yang akan dibangkitkan pada setiap data minor
   b. Memberikan bobot r_i
   c. Melakukan normalisasi pada bobot r_i
   d. Mereplikasi setiap data minor sebanyak g_i
6. Melakukan penanganan data *imbalance* menggunakan metode ADASYN-kNN, dengan mengulang langkah-langkah 5a hingga 5c, kemudian dilakukan pembentukan data sintetis dengan mencari mayoritas tetangga terdekat pada setiap data minor kemudian dilakukan replikasi sebanyak g_i.
7. Melakukan analisis klasifikasi faktor-faktor yang diduga menyebabkan kasus *Stunting* pada balita menggunakan *classification and regression tree* (CART)
   a. Melakukan pembentukan pohon klasifikasi optimal
   b. Melakukan pemangkasan pohon klasifikasi

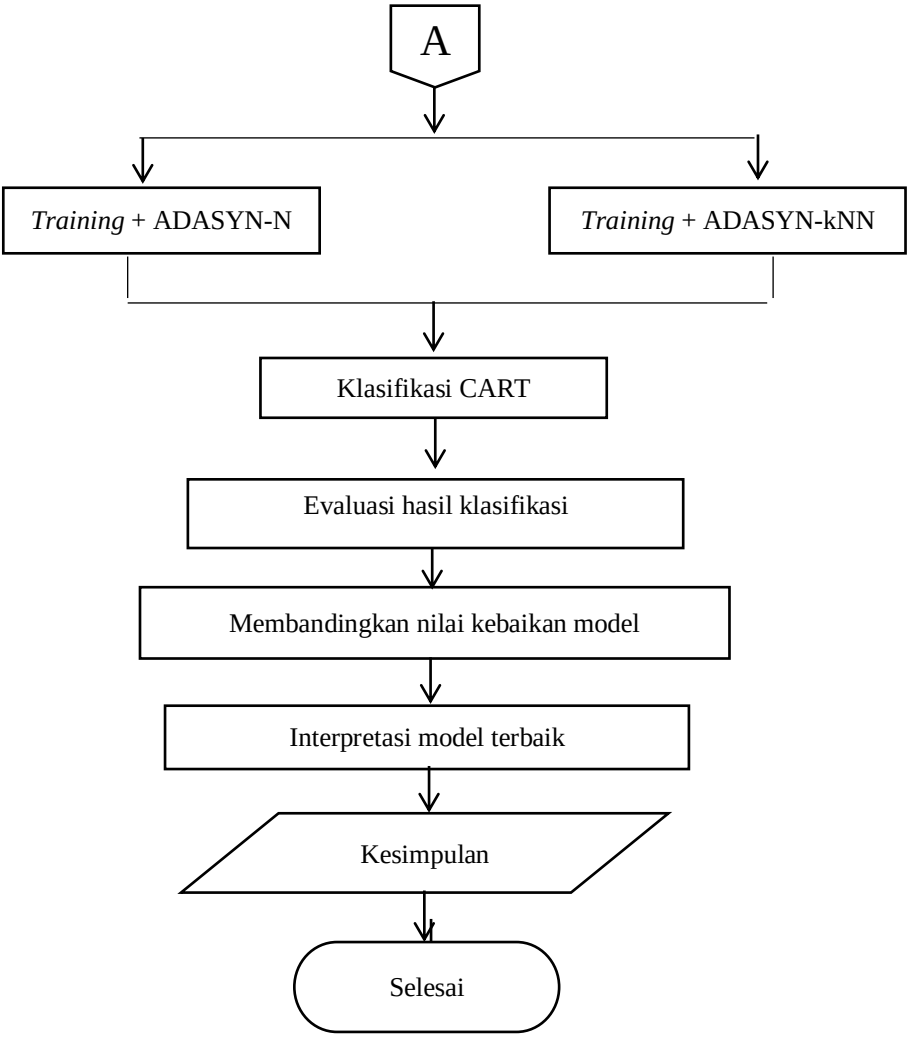
c. Melakukan pemilihan pohon terbaik
d. Melakukan evaluasi hasil klasifikasi menggunakan rata-rata nilai AUC pada data *testing*.

8. Membandingkan nilai rata-rata AUC dari hasil klasifikasi menggunakan berbagai metode.
9. Menginterpretasikan hasil klasifikasi dengan kinerja klasifikasi terbaik.
10. Menarik kesimpulan dan saran.

3.5 Diagram Alir

Struktur Diagram alir dari langkah analisis pada penelitian ini disajikan pada Gambar 3.1.





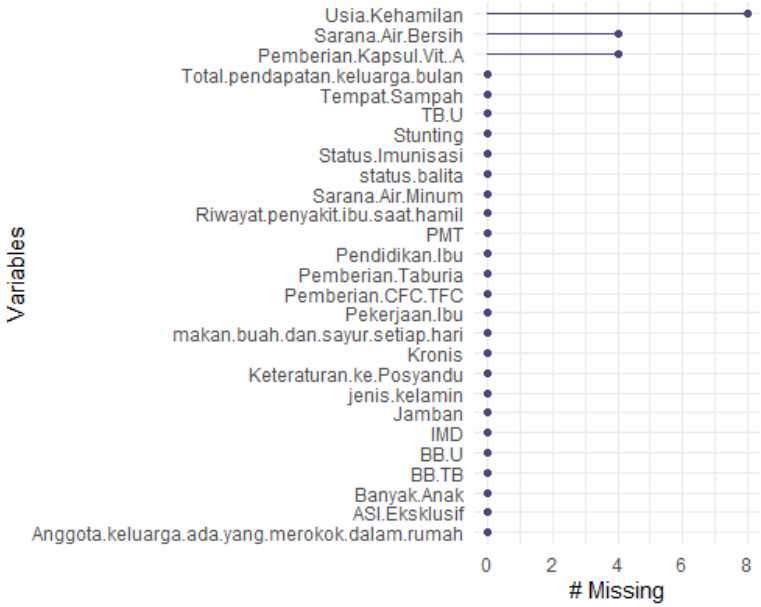
Gambar 3. 1 Diagram Alir

BAB IV
ANALISIS DAN PEMBAHASAN

Pada penelitian ini akan dilakukan klasifikasi pada balita di Kota Surabaya terkait kasus *stunting*. Sebelumnya pada tahun 2019, di Kota Surabaya dilakukan pemeriksaan balita oleh Kader Posyandu yang juga bekerjasama dengan Puskesmas setempat di setiap kelurahan di Kota Surabaya untuk mengetahui balita-balita yang diduga mengalami kasus *stunting*, kemudian pada Agustus 2019 hingga Februari 2020 dilakukan pemeriksaan kembali oleh tenaga kesehatan pada 4754 balita yang diduga mengalami stunting. Setelah dilakukan pemeriksaan kembali, dari 4754 balita tidak semua balita yang diduga stunting mengalami kasus stunting, namun mayoritas dari balita yang diperiksa mengalami stunting sehingga dilakukan klasifikasi menggunakan *Classification and Regression Trees* (CART) dikarenakan data yang cukup besar dan metode ini dapat digunakan pada data yang berjumlah besar. Klasifikasi CART dilakukan dengan metode penanganan data *imbalance Adaptive Synthetic Nominal* (ADASYN-N) dan *Adaptive Synthetic k-Nearest Neighbor* (ADASYN-kNN) dikarenakan data balita *stunting* lebih banyak dari data balita normal (data tidak seimbang). Klasifikasi ini digunakan untuk mengetahui faktor-faktor yang mengindikasikan terjadinya *stunting* pada balita. Pada penelitian ini juga akan dibandingkan metode penanganan data *imbalance* terbaik dengan membandingkan nilai *Area Under Curve* (AUC).

4.1 *Preprocessing Data*

Preprocessing data ini dilakukan sebelum penanganan data *imbalance*, *preprocessing data* merupakan proses untuk membuat data mentah menjadi data yang siap untuk diolah, pada penelitian ini *preprocessing data* yang dilakukan adalah *missing value*. Hasil dari *missing value* dari tiap variabel ditunjukkan pada Gambar 4.1.



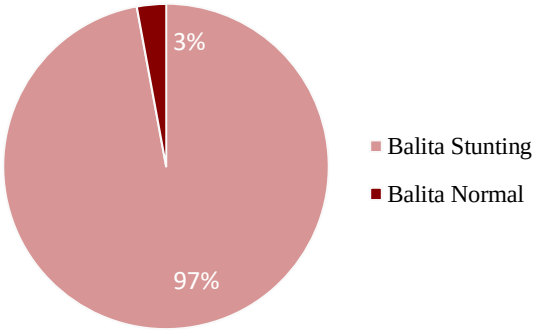
Gambar 4. 1 *Output Missing Value*

Gambar 4.1 menunjukkan bahwa pada variabel usia kehamilan terdapat 8 *missing value*, sarana air bersih terdapat 4 *missing value*, dan pada variabel pemberian kapsul vitamin A terdapat 4 *missing value*. Kemudian untuk mengatasi *missing value* tersebut dilakukan imputasi. Imputasi dilakukan dengan menggunakan nilai modus. Pada variabel usia kehamilan diimputasi dengan nilai 1 dimana nilai 1 adalah kategori ibu dari balita yang memiliki usia kehamilan ≥37 minggu saat mengandung balita tersebut. Kemudian pada variabel sarana air bersih diimputasi dengan nilai 2 yaitu kategori bahwa sarana air bersih rumah yang dihuni balita telah memenuhi syarat, serta pada variabel pemberian kapsul vitamin A diimputasi dengan nilai 1 dimana nilai 1 merupakan kategori telah dilakukan pemberian kapsul vitamin A pada balita. Setelah dilakukan

penanganan terhadap *missing value*, maka dapat dilanjutkan ke analisis berikutnya yaitu eksplorasi data.

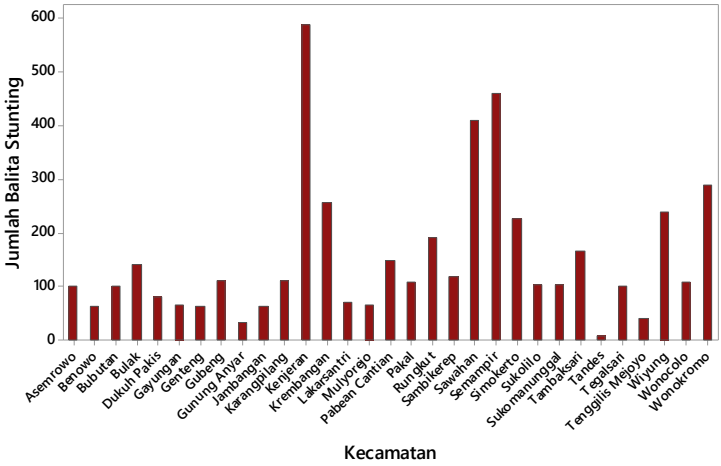
4.2 Eksplorasi Data

Setelah dilakukan *preprocessing data* kemudian dilakukan eksplorasi data untuk mengetahui gambaran dari data. Data dari penelitian ini merupakan data balita dari 31 kecamatan di Kota Surabaya terkait kasus stunting. Data terdiri dari 27 variabel dan merupakan data kategorik. Data dikelompokkan menjadi dua kategori yaitu kategori balita normal dan kategori balita stunting.



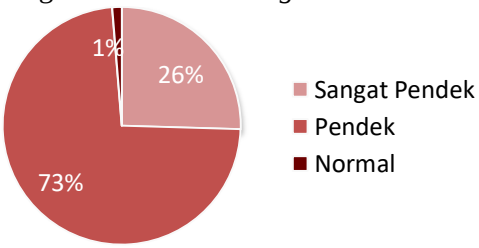
Gambar 4. 2 Persentase Balita Terkait Kasus Stunting

Pada Gambar 4.2 terlihat bahwa dari 4754 balita di Kota Surabaya yang diduga mengalami stunting, sebesar 97% atau sebanyak 4614 balita terkonfirmasi mengalami stunting saat diperiksa oleh tenaga kesehatan, sedangkan sisanya sebesar 3% atau sebanyak 140 balita tidak mengalami stunting (balita normal). Kemudian untuk jumlah balita stunting di setiap kecamatan terdapat pada Gambar 4.3.



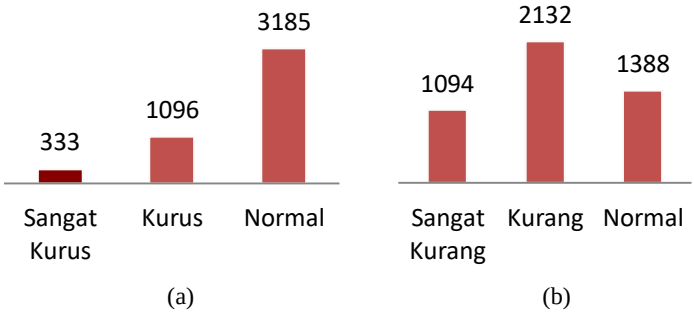
Gambar 4. 3 *BarChart* Balita *Stunting* di Setiap Kecamatan di Kota Surabaya

Gambar 4.3 menunjukkan bahwa Kecamatan Kenjeran memiliki jumlah kasus *stunting* tertinggi yaitu sebesar 588 balita di Kecamatan Kenjeran mengalami kasus *stunting* diikuti dengan Kecamatan Semampir yaitu sebanyak 459 balita yang mengalami kasus *stunting*, sedangkan kecamatan dengan kasus *stunting* terendah berada di Kecamatan Tandes yaitu hanya sebanyak 9 balita yang mengalami kasus *stunting*.



Gambar 4. 4 Persentase Kategori Balita Stunting Berdasarkan TB/U

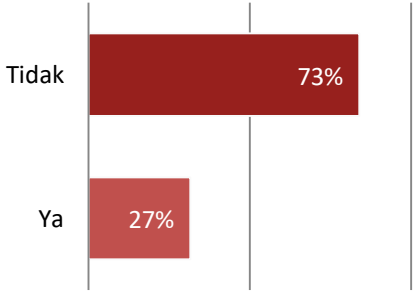
Gambar 4.4 menunjukkan bahwa dari 4614 balita yang mengalami stunting, sebanyak 73% balita memiliki tinggi badan yang dikategorikan sangat pendek apabila dibandingkan dengan umurnya, 26% balita memiliki tinggi badan yang dikategorikan pendek dibandingkan dengan umur, dan balita yang memiliki tinggi badan normal dibandingkan umur hanya 1%.



Gambar 4. 5 *Bar Chart* Kategori Balita *Stunting* Berdasarkan (a) BB/TB dan (b) BB/U

Selanjutnya berdasarkan kategori berat badan dibanding tinggi badan pada Gambar 4.5 dapat diketahui bahwa sebanyak 333 balita *stunting* atau sebesar 7% balita *stunting* dikategorikan sangat kurus, 1096 balita stunting atau 24% balita *stunting* dikategorikan kurus, dan sebanyak 3185 balita *stunting* atau sebesar 69% balita *stunting* memiliki berat badan yang normal apabila dibandingkan dengan tinggi badannya. Balita yang masuk pada kategori normal cukup banyak dibanding balita yang masuk pada kategori sangat kurus dan kurus berdasarkan kategori berat badan dibandingkan tinggi badan. Namun apabila berdasarkan kategori berat badan dibandingkan dengan umur, sebanyak 1094 balita *stunting* atau sebesar 24% balita *stunting* dikategorikan memiliki berat badan yang sangat kurang dibandingkan umur, 2132 balita *stunting* atau sebesar 46% balita *stunting* dikategori memiliki berat badan yang kurang dibandingkan umur, dan

sebanyak 1388 atau sebesar 30% balita *stunting* memiliki berat badan yang normal dibandingkan umur. Masih cukup tingginya balita *stunting* yang berada di luar kategori normal menunjukkan bahwa pemenuhan gizi pada balita dan perilaku hidup sehat tersebut masih kurang. Salah satu usaha dalam pemenuhan gizi adalah dengan selalu makan makanan yang bergizi. Pada penelitian ini terdapat variabel makan buah dan sayur setiap hari bagi balita sebagai salah satu variabel untuk melihat perilaku hidup sehat dan pemenuhan gizi.

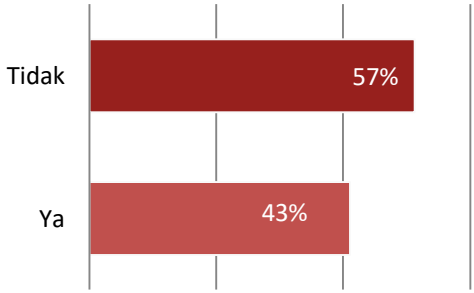


Gambar 4. 6 Persentase Balita *Stunting* Makan Buah dan Sayur Setiap Hari

Persentase balita *stunting* makan buah dan sayur setiap hari ditunjukkan pada Gambar 4.6, pada gambar tersebut terlihat bahwa sebesar 73% balita *stunting* tidak makan buah dan sayur setiap hari dan 27% balita *stunting* makan buah dan sayur setiap hari. Dari 31 kecamatan, hanya satu kecamatan dengan persentase balita *stunting* yang memiliki kebiasaan makan buah dan sayur setiap hari diatas 50 persen, yaitu Kecamatan Genteng. Pada Kecamatan Genteng, dari 65 balita *stunting* sebanyak 34 balita *stunting* telah memiliki kebiasaan makan buah dan sayur setiap hari. Sedangkan balita *stunting* di 30 kecamatan yang lain memiliki persentase kebiasaan makan buah dan sayur setiap hari dibawah 50 persen. Hal ini menunjukkan bahwa kebiasaan memakan makanan sehat dan bergizi seperti buah dan sayur

setiap hari oleh balita *stunting* di Kota Surabaya masih rendah sehingga peningkatan terhadap kesadaran memakan makanan bergizi harus lebih ditingkatkan.

Selain pemenuhan makanan pokok yang harus memenuhi kebutuhan gizi balita, pemerintah mengembangkan program Pemberian Makanan Tambahan (PMT), makanan tambahan tersebut dikonsumsi oleh balita yang mengalami gizi buruk dan sebagai makanan tambahan sehari-hari bukan pengganti makanan pokok. Makanan tambahan dapat berupa snack atau makanan ringan yang memiliki mutu yang baik dan telah memenuhi standar gizi yang diperlukan oleh balita sebagai makanan pendamping.

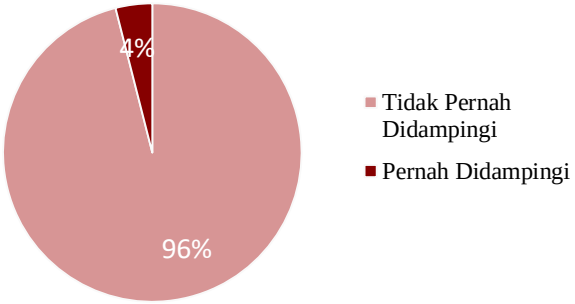


Gambar 4. 7 Persentase Balita Stunting yang Memperoleh PMT

Pada Gambar 4.7, terlihat bahwa dari 4614 balita yang mengalami *stunting*, sebanyak 43% balita sudah memperoleh pemberian makanan tambahan, namun 57% balita stunting belum mendapatkan pemberian makanan tambahan. Sehingga dapat disimpulkan bahwa masih cukup banyak balita stunting yang belum mendapatkan Pemberian Makanan Tambahan (PMT).

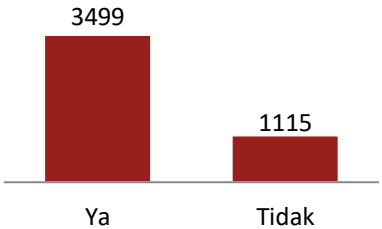
Selanjutnya dalam upaya penanganan terhadap balita kekurangan gizi atau gizi buruk dimana balita yang kekurangan gizi atau gizi buruk memiliki beresiko mengalami stunting, pemerintah Dinas Kesehatan Kota Surabaya melakukan

perekrutan tenaga gizi yang kemudian ditempatkan di wilayah puskesmas tertentu untuk mendampingi balita yang mengalami kekurangan gizi atau gizi buruk.



Gambar 4. 8 Persentase Pendampingan Balita Stunting

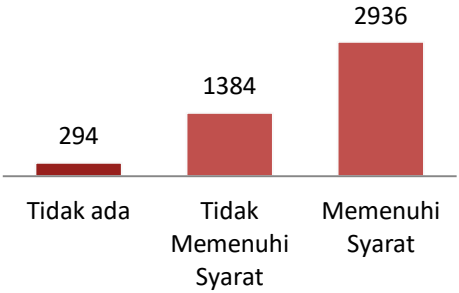
Gambar 4.8 menunjukkan bahwa hanya 4% atau sebanyak 183 balita stunting yang pernah didampingi oleh tenaga gizi, dan sisanya yaitu sebesar 96% atau sebanyak 4431 balita *stunting* belum pernah didampingi oleh tenaga gizi. Hal ini menunjukkan bahwa perlu adanya peningkatan pendampingan tenaga gizi terhadap balita yang kekurangan gizi ataupun gizi buruk dalam rangka penanganan gizi buruk yang dapat mengarah ke stunting dikarenakan balita yang mengalami gizi buruk juga memiliki resiko mengalami *stunting*. Selanjutnya analisis deskriptif dilakukan pada variabel kelengkapan imunisasi dasar dari balita stunting yang ditunjukkan pada Gambar 4.9.



Gambar 4. 9 Bar Chart Kelengkapan Imunisasi Balita Stunting

Gambar 4.9 menunjukkan bahwa sebanyak 3499 balita stunting telah melakukan imunisasi dasar dengan lengkap, sedangkan sebanyak 1115 balita stunting belum melakukan imunisasi dasar dengan lengkap.

Selanjutnya, pola hidup bersih dan sehat juga perlu dilakukan untuk mencegah tumbuhnya bakteri dan virus yang menyebabkan penyakit. Salah satu upaya untuk hidup bersih dan sehat adalah dengan memiliki sarana air bersih yang layak.



Gambar 4. 10 Bar Chart Keluarga Balita Stunting yang Memilliki Sarana Air Bersih

Sebanyak 2936 keluarga dari balita *stunting* telah memiliki sarana air bersih yang memenuhi syarat, yaitu sarana air bersih miliki sendiri dan telah memenuhi syarat kesehatan, kemudian sebanyak 1384 keluarga memiliki sarana air bersih yang tidak memenuhi syarat, yaitu sarana air bersih bukan milik sendiri dan memenuhi syarat kesehatan ataupun sarana air bersih milik sendiri namun tidak memenuhi syarat kesehatan, sedangkan 294 keluarga belum memiliki sarana air bersih.

4.3 Analisis Klasifikasi Balita Terkait Kasus Stunting Menggunakan CART dengan ADASYN-N

Pada penelitian ini dilakukan klasifikasi balita di Kota Surabaya terkait kasus *stunting* dengan menggunakan metode *Classification and Regression Trees* (CART). Pada penelitian ini

data merupakan data *imbalance* dikarenakan jumlah balita *stunting* yang jauh lebih besar dibanding jumlah balita normal. Metode penanganan data *imbalance* pertama yang digunakan pada penelitian ini adalah metode *Adaptive Synthetic Nominal* (ADASYN-N). ADASYN-N merupakan pengembangan dari metode ADASYN yang digunakan pada data nominal, ADASYN-N menggunakan versi modifikasi dari *Value Difference Metric* (VDM) untuk menghitung jarak antar kategori. Pada penelitian ini *degree of class imbalance* (*d*) adalah sebesar 0,03 dimana $d \in [0,1]$, sehingga data merupakan data *imbalance*. Data yang digunakan pada analisis dengan menggunakan ADASYN-N merupakan data *training*, data *training* diperoleh dari pembagian data *training* dan *testing* secara acak dengan persentase data *training* yang ditentukan penulis sebesar 80% dan data *testing* sebesar 20%. Pembagian data *training* dan data *testing* diulang sebanyak 5 kali secara acak sehingga terdapat 5 partisi yang berbeda. Analisis dilakukan pada 5 partisi yang memiliki data *training* dengan pengacakan yang berbeda untuk mengetahui hasil prediksi terbaik diantara 5 partisi tersebut. Selanjutnya dilakukan ilustrasi analisis ADASYN-N pada CART. Berikut ilustrasi dilakukan pada partisi 1 yang terdiri dari 3804 data *training*, 1 variabel respon, dan 26 variabel prediktor.

1. Menentukan jumlah *instance* data sintetis yang perlu digeneralisasi untuk kelas minoritas (*G*)

Sebelum menentukan nilai *G*, dari 3804 data *training* diketahui bahwa jumlah *instance* kelas minoritas sebesar 112 data dan jumlah *instance* kelas mayoritas adalah sebesar 3692 data. Kemudian dilakukan perhitungan nilai *G* sebagai berikut.

$$G = (3692 - 112) \times 1$$
$$= 3580$$

Nilai 1 merupakan nilai β, dimana $\beta = 1$ menunjukkan data yang sepenuhnya seimbang setelah digeneralisasi. Berdasarkan perhitungan, maka data sintetis yang perlu dibangkitkan untuk kelas minoritas adalah sebesar 3580 data.

2. Menentukan k-*nearest neighbors*, rasio r_i, dan $\hat{r}_i$

Setelah menentukan nilai *G* kemudian menentukan k-*nearest neighbors*, sebagai ilustrasi apabila dilakukan pengambilan pada *instance* ke-2693, maka tetangga terdekat dari *instance* ke-2693 adalah 2, 3, 4, 5, 6, 7, 10, 11, 12, dan 14. Dari 10 tetangga terdekat tersebut tidak ada yang merupakan kelas minoritas, sehingga 10 tetangga terdekat tersebut merupakan kelas mayoritas. Kemudian dilakukan perhitungan rasio r_i dari *instance* ke-2693.

$$r_1 = \frac{\Delta_1}{k} = \frac{10}{10} = 1$$

Perhitungan r_i dilakukan pada setiap *instance* minoritas, sehingga setiap instance minoritas memiliki rasio r_i, kemudian dilakukan normalisasi pada r_i sebagai berikut.

$$\hat{r}_1 = \frac{r_1}{\sum_{i=1}^{m_s} r_i} = \frac{1}{112} = 0,00893$$

Berdasarkan perhitungan diketahui bahwa nilai $\hat{r}_1$ adalah sebesar 0,00893. Selanjutnya dilakukan perhitungan jumlah *instance* sintetis yang perlu dihasilkan.

3. Menghitung jumlah *instance* sintetis yang perlu dihasilkan bagi setiap *instance* minoritas

Setelah mengetahui jumlah *instance* data sintetis yang perlu digeneralisasi untuk kelas minoritas serta nilai $\hat{r}_1$, kemudian

dilakukan perhitungan untuk mengetahui jumlah *instance* sintetis yang perlu dihasilkan. Jumlah *instance* yang perlu dihasilkan bagi setiap *instance* minoritas diperoleh dari banyak replikasi sebanyak g_i. Berikut merupakan perhitungan jumlah *instance* sintetis yang perlu dihasilkan pada *instance* ke-2693.

$$\begin{aligned} g_1 &= \hat{r}_1 \times G \\ &= 0,00893 \times 3580 \\ &= 31,97 \approx 32 \end{aligned}$$

Berdasarkan perhitungan, maka *instance* 2693 akan direplikasi sebanyak 32 kali. Perhitungan ini dilakukan pada setiap *instance* minoritas sehingga kemudian diperoleh data yang seimbang.

Setelah diperoleh data yang seimbang kemudian dapat dilanjutkan ke analisis berikutnya yaitu klasifikasi dengan menggunakan *Classification and Regression Trees* (CART). Pada penelitian ini akan dihasilkan pohon klasifikasi dikarenakan data yang digunakan merupakan data berskala kategorik. Terdapat tiga tahapan dalam analisis menggunakan metode CART yaitu pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi, serta penentuan pohon klasifikasi yang optimum. Berikut merupakan ilustrasi analisis klasifikasi dengan menggunakan metode CART pada partisi 1.

Langkah pertama dalam klasifikasi dengan menggunakan metode CART adalah pembentukan pohon klasifikasi (*Tree Building Process*). Pada pembentukan pohon klasifikasi, proses awal yang dilakukan adalah melakukan proses pemisahan simpul pada setiap simpul yang untuk mendapatkan pemilah yang mampu menghasilkan simpul dengan tingkat kehomogenan nilai peubah respon paling tinggi. Proses pemilahan dilakukan dengan menggunakan indeks gini. Berikut merupakan ilustrasi

perhitungan indeks gini pada variabel anggota keluarga yang merokok di dalam rumah.

Tabel 4. 1 Pemilahan pada Simpul Anggota Keluarga Merokok

Anggota Keluarga Merokok dalam Rumah	*Stunting*		Total
	Ya	Tidak	
Merokok	2178	2871	5049
Tidak Merokok	1514	825	2339

Selanjutnya dilakukan perhitungan indeks gini pada simpul kiri dan simpul kanan sebagai berikut.

$$r(t_L) = 2 \times \left(\frac{1514}{2339} \times \frac{825}{2339} \right) = 0,4566$$

$$r(t_R) = 2 \times \left(\frac{2178}{5049} \times \frac{2871}{5049} \right) = 0,4906$$

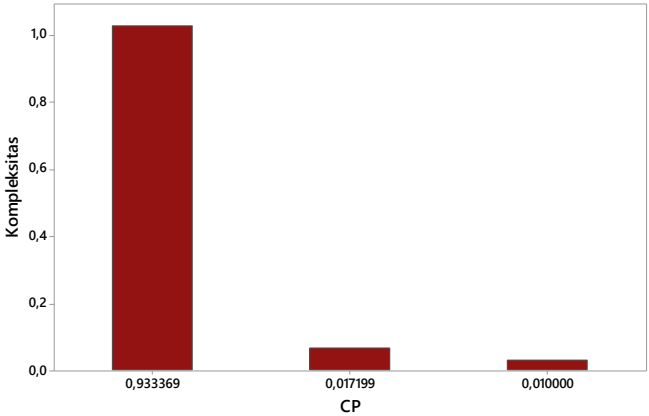
Setelah dilakukan perhitungan indeks gini pada simpul kanan dan kiri kemudian dilanjutkan dengan menentukan kriteria *goodness of spli*t yang merupakan suatu evaluasi pemilahan yang dilakukan oleh pemilah pada suatu simpul. Berikut merupakan perhitungan *goodness of spli*t pada variabel anggota keluarga yang merokok di dalam rumah.

$$\phi(s,t) = 0,5 - \left(\frac{2339}{7388} \right) \times 0,4566 - \left(\frac{5049}{7388} \right) \times 0,4906 = 0,0202$$

Pemilah yang menghasilkan nilai *goodness of split* tinggi merupakan pemilah terbaik karena dapat menghasilkan heterogenitas lebih tinggi. Kemudian dibentuk pohon dari pemilah-pemilah berdasarkan nilai *goodness of split* tertinggi ke terendah. Setelah itu dari simpul-simpul yang ada dilakukan penentuan simpul terminal dimana simpul t dikatakan sebagai simpul terminal apabila tidak terdapat penurunan heterogenitas

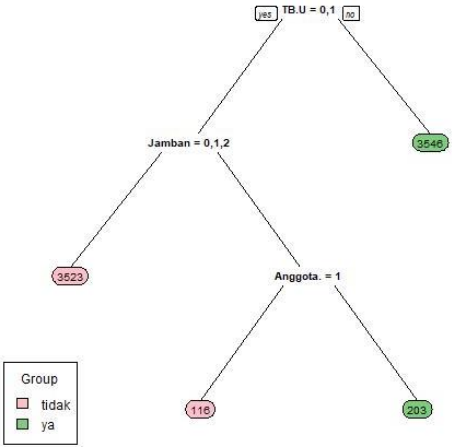
yang signifikan atau hanya terdapat satu pengamatan di setiap simpul anak. Kemudian dari simpul terminal yang ada dilakukan penentuan label kelas yaitu simpul terminal akan terlabel *stunting* atau normal berdasarkan aturan jumlah terbanyak di simpul terminal tersebut dengan menggunakan rumus 2.4.

Setelah melakukan pembentukan pohon klasifikasi, langkah selanjutnya adalah melakukan pemangkasan pohon klasifikasi untuk menghindari terjadinya *overfitting*. Pemangkasan pohon didasarkan pada suatu penilaian ukuran sebuah pohon dengan menggunakan ukuran *cost complexity minimum*. Pada klasifikasi balita di Kota Surabaya terkait kasus stunting, diperoleh nilai *cp* yang optimal sebesar 0,01 seperti pada Gambar 4.11.



Gambar 4. 11 *Complexity Parameter* terhadap Data *Imbalance*

Setelah diperoleh nilai *cp* yang optimum, kemudian dilakukan pembentukan pohon klasifikasi menggunakan *cp* optimum yang ditunjukkan pada Gambar 4.12.



Gambar 4. 12 Pohon Klasifikasi Partisi 1

Gambar 4.12 menunjukkan bahwa tinggi badan balita dibandingkan umur merupakan variabel utama yang mengindikasikan terjadinya kasus *stunting* pada balita, kemudian dari variabel tinggi badan balita sebagai simpul induk terbentuk simpul terminal yaitu variabel jamban yang digunakan, selanjutnya dari variabel jamban yang digunakan terbentuk simpul terminal berupa variabel anggota keluarga tidak ada yang merokok di dalam rumah. Variabel-variabel tersebut merupakan variabel yang mengindikasikan terjadinya kasus *stunting* pada balita. Sedangkan variabel lainnya yang digunakan pada penelitian ini tidak mengindikasikan terjadinya kasus *stunting* balita pada penelitian ini. Setelah dilakukan klasifikasi, kemudian dibentuk *confusion matrix* seperti pada Tabel 4.2.

Tabel 4. 2 *Confusion Matrix Adasyn-N* Partisi 1

Aktual	Prediksi	
	Positive	Negative
Positive	3606	86
Negative	33	3663

Tabel 4.2 menunjukkan bahwa pada partisi 1, model mengklasifikasikan secara benar sebesar 3663 data sebagai negatif dan 3606 data sebagai positif. Selain itu model salah dalam memprediksi 86 data ke dalam data negatif yang seharusnya positif (*false negative*), serta salah dalam memprediksi 33 data ke dalam data positif yang seharusnya negatif (*false positive*).

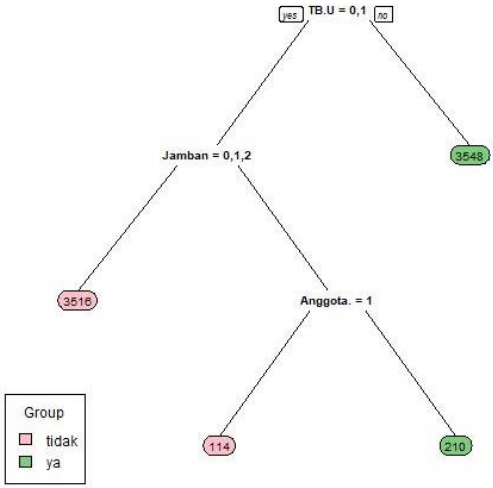
Selanjutnya untuk nilai AUC dari 5 partisi data *imbalance* yang telah ditangani dengan metode ADASYN-N ditunjukkan pada Tabel 4.3.

Tabel 4. 3 Nilai AUC Hasil Klasifikasi CART dengan ADASYN-N

Partisi	**AUC *Training***	**AUC *Testing***
1	0,9838889	0,9685854
2	0,9832118	0,9712969
3	0,9826701	**0,9734661**
4	0,9837535	0,9691277
5	0,9842952	0,9669585

Rata-rata nilai AUC yang dihasilkan dari klasifikasi CART dengan ADASYN-N menunjukkan nilai yang cukup tinggi seperti yang ditunjukkan pada Tabel 4.3. Dari 5 partisi, AUC *training* tertinggi yaitu sebesar 98,43% terdapat pada partisi 5 yang berarti bahwa partisi 5 memiliki model yang terbaik. Kemudian untuk AUC *testing* tertinggi yaitu sebesar 97,35% yang terdapat pada partisi 3 yang berarti hasil prediksi dari partisi 3 merupakan hasil prediksi terbaik. Sehingga selanjutnya dibentuk pohon klasifikasi dari partisi 3.

Pohon klasifikasi dari partisi 3 dimana partisi 3 merupakan partisi dengan AUC *testing* tertinggi ditunjukkan pada Gambar 4.13.



Gambar 4. 13 Pohon Klasifikasi Terbaik dengan ADASYN-N

Gambar 4.13 menunjukkan bahwa variabel tinggi badan balita dibandingkan umur menjadi variabel utama dalam mengklasifikasikan balita terkait kasus *stunting* di Kota Surabaya. Selanjutnya untuk mengetahui karakteristik balita yang digolongkan *stunting*, Tabel 4.4 menunjukkan *gain nodes* dari pohon keputusan terbaik pada Gambar 4.13.

Tabel 4. 4 *Gains Nodes* Balita Mengalami Kasus *Stunting* (ADASYN-N)

Simpul Terminal	Karateristik Balita yang Mengalami *Stunting*
1.	a. Balita yang memiliki tinggi badan dibandingkan umur masuk dalam kategori pendek, atau balita yang memiliki tinggi badan dibandingkan umur masuk dalam kategori sangat pendek, b. Rumah balita tidak memiliki jamban /BAB sembarangan, dan c. Balita tidak memiliki anggota keluarga yang merokok di dalam rumah.
2	a. Balita memiliki tinggi badan dibandingkan umur normal

Pohon klasifikasi tersebut memiliki dua simpul terminal yang memiliki kategori *stunting*. Pada simpul terminal pertama, balita yang mengalami kasus *stunting* merupakan balita yang memiliki tinggi badan dibandingkan umur yang masuk dalam kategori sangat pendek dan pendek, kemudian rumah dari balita tersebut tidak memiliki jamban sehat dan buang air besar secara sembarangan, selain itu tidak terdapat anggota keluarga balita yang merokok di dalam rumah. Tidak adanya anggota keluarga yang merokok di dalam rumah dapat mengindikasikan terjadinya *stunting* kemungkinan dikarenakan pada kasus ini faktor adanya keluarga merokok di dalam rumah tidak secara langsung mengindakasi terjadinya *stunting*, namun terdapat faktor lain terkait anggota keluarga yang merokok di dalam rumah, seperti faktor ekonomi dimana penghasilan yang ada digunakan untuk biaya membeli rokok sehingga mengurangi biaya untuk kebutuhan bahan pangan yang menyebabkan anak kekurangan makanan bergizi, atau faktor kebersihan dimana setelah anggota keluarga merokok di luar rumah tidak segera membersihkan badan dan sisa-sisa asap rokok yang menempel pada pakaian terhirup oleh anak. Anak yang terus menerus menghirup asap rokok dapat mengalami penyakit pernapasan kronis dimana penyakit kronis merupakan salah satu faktor protektif malnutrisi pada balita, terjadinya malnutrisi pada balita kemudian mempertinggi resiko terjadinya *stunting*. Namun masih perlu dilakukan penelitian lebih lanjut terkait dampak kebiasaan merokok anggota keluarga terhadap kesehatan anak yang dapat mengarah pada penyakit infeksi dan gangguan kesehatan sehingga meningkatkan resiko terjadinya *stunting*. Selanjutnya pada simpul terminal kedua, balita yang memiliki tinggi badan normalpun dapat mengalami kasus *stunting*, hal ini dapat dilihat pada Gambar 4.4, meskipun mayoritas balita yang mengalami *stunting* memiliki tinggi badan dibanding umur dalam kategori sangat pendek ataupun pendek, namun sebanyak 1% dari balita

stunting atau sebanyak 60 balita *stunting* memiliki tinggi badan yang normal, sehingga balita dengan tinggi badan normalpun bisa beresiko terkena *stunting* apabila memiliki gizi yang buruk.

4.4 Analisis Klasifikasi Balita Terkait Kasus *Stunting* Menggunakan CART dengan ADASYN-kNN

Metode penanganan data *imbalance* kedua yang digunakan untuk analisis klasifikasi dengan menggunakan metode CART adalah metode ADASYN-kNN. Metode ADASYN-kNN memiliki algoritma yang hampir sama dengan ADASYN-N, perbedaannya terletak pada proses pembentukan *instance* baru dimana pada ADASYN-kNN diperoleh *instance* baru dengan atribut berdasarkan pada *majority voting* dari tetangga terdekat *instance* minoritas. Berikut merupakan perhitungan ADASYN-kNN pada klasifikasi dengan menggunakan CART pada partisi 1.
Pada analisis ADASYN-kNN pada partisi 1, untuk langkah penentuan jumlah *instance* data sintetis yang perlu digeneralisasi untuk kelas minoritas sama seperti pada ADASYN-N sehingga penentuan jumlah *instance* data sintetis yang perlu digeneralisasi untuk kelas minoritas yaitu sebesar 3580 *instance*. Kemudian dilakukan penentuan tetangga terdekat dari *instance* dan dilakukan perhitungan rasio r_i dan dilakukan normalisasi ($\hat{r}_i$). Setelah dilakukan proses normalisasi langkah selanjutnya yang membedakan dengan ADASYN-N adalah membentuk *instance* baru dengan mencari tetangga terdekat, dimana hal ini dilakukan pada seluruh *instance* minoritas pada partisi 1. *Instance* baru terbentuk dari *majority voting* 10 tetangga terdekat. Setelah terbentuk *instance* baru, maka dilakukan replikasi *instance* baru tersebut sebanyak g_i, sehingga diperoleh data *training* baru sebanyak 7388 data.

Perhitungan dilakukan pada seluruh partisi. Selanjutnya setelah diperoleh data yang seimbang, maka dilakukan klasifikasi dengan menggunakan metode CART. Nilai AUC dari klasifikasi

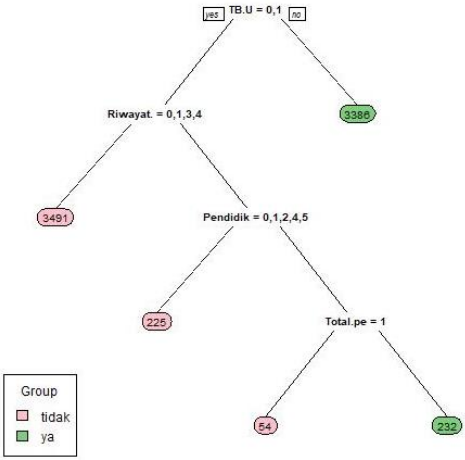
CART dengan penanganan data *imbalance* ADASYN-kNN ditunjukkan pada Tabel 4.5.

Tabel 4. 5 Nilai AUC Hasil Klasifikasi CART dengan ADASYN-kNN

Partisi	AUC *Training*	AUC *Testing*
1	0,9681988	0,9561511
2	**0,9841555**	0,9290749
3	0,9769921	0,9556089
4	0,974559	**0,9561512**
5	0,9791542	0,9491013

Pada Tabel 4.5, AUC *training* tertinggi sebesar 98,42% pada partisi 2 yang berarti partisi 2 memiliki model terbaik dan AUC *training* terendah sebesar 96,82% pada partisi 1. Sedangkan AUC *testing* tertinggi yaitu sebesar 95,62% pada partisi 4 yang berarti partisi 4 memiliki hasil prediksi terbaik dan AUC *testing* terendah sebesar 92,91% pada partisi 2.

Partisi 4 memiliki nilai AUC *testing* terbaik yang berarti memiliki hasil prediksi terbaik sehingga kemudian dibuat pohon klasifikasi dari partisi 4. Pohon klasifikasi ditampilkan pada Gambar 4.14.



Gambar 4. 14 Pohon Klasifikasi Terbaik dengan ADASYN-kNN

Gambar 4.14 menunjukkan bahwa tinggi badan balita dibandingkan umur menjadi variabel pertama dalam mengklasifikasikan balita terkait kasus *stunting*. Selain variabel tinggi badan dibandingkan umur, variabel riwayat penyakit ibu hamil, pendidikan terakhir ibu, serta total pendapatan keluarga. *Gain nodes* dari pohon keputusan terbaik hasil klasifikasi CART dengan ADASYN-kNN ditunjukkan pada Tabel 4.6.

Tabel 4. 6 *Gains Nodes* Balita Mengalami Kasus *Stunting* (ADASYN-kNN)

Simpul Terminal	Karateristik Balita yang Mengalami *Stunting*
1.	a. Balita yang memiliki tinggi badan dibandingkan umur yang normal, balita yang memiliki tinggi badan dibandingkan umur yang masuk kategori sangat pendek, b. Ibu balita saat hamil memiliki riwayat penyakit diabetes mellitus dan penyakit lainnya selain hipertensi, pre eklamsia, dan ketuban pecah dini, c. Ibu balita memiliki pendidikan terakhir SMP d. Total pendapatan keluarga <UMK
2	a. Balita yang memiliki tinggi badan dibandingkan umur yang normal

Pada simpul terminal kategori *stunting* pertama, balita yang mengalami *stunting* merupakan balita yang memiliki tinggi badan dibandingkan umur masuk dalam kategori sangat pendek atau masuk dalam kategori pendek, kemudian ibu dari balita saat hamil memiliki riwayat penyakit diabetes mellitus atau riwayat penyakit lainnya selain hipertensi, pre eklamsia, dan ketuban pecah dini, selain itu ibu balita memiliki pendidikan terakhir SMP dan total pendapatan keluarga balita <UMK per bulan. Pada simpul terminal kedua, balita yang memiliki tinggi badan normal juga merupakan karakteristik balita yang mengalami *stunting*.

4.5 Perbandingan Metode

Klasifikasi dengan menggunakan metode CART dengan dua metode penanganan data *imbalance*, yaitu metode ADASYN-

N dan ADASYN-kNN menghasilkan nilai AUC yang ditunjukkan pada Tabel 4.7.

Tabel 4. 7 Perbandingan Nilai AUC *Testing* pada Klasifikasi CART

Partisi	AUC dengan ADASYN-N	AUC dengan ADASYN-kNN
1	0,9685854	0,9561511
2	0,9712969	0,9290749
3	**0,9734661**	0,9556089
4	0,9691277	**0,9561512**
5	0,9669585	0,9491013

AUC *testing* pada penelitian ini digunakan untuk mengetahui seberapa baik hasil prediksi dari model klasifikasi. Tabel 4.7 menunjukkan bahwa nilai AUC *testing* tertinggi dengan menggunakan metode ADASYN-N adalah sebesar 97,35% pada partisi 3 yang berarti partisi 3 memiliki hasil prediksi terbaik. Sedangkan hasil prediksi terbaik dengan menggunakan penanganan *imbalance* ADASYN-kNN ditunjukkan dengan nilai AUC *testing* tertinggi sebesar 95,62% pada partisi 4. Nilai AUC terbaik dalam klasifikasi CART pada penelitian ini yaitu dengan menggunakan penanganan data *imbalance* ADASYN-N pada partisi 3 dengan nilai AUC sebesar 97,35%.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan, maka kesimpulan yang diperoleh dari penelitian ini adalah.

1. Kecamatan Kenjeran memiliki jumlah kasus *stunting* tertinggi dan kasus terendah berada di Kecamatan Tandes. Untuk kesadaran pemenuhan gizi, kesadaran untuk memakan buah dan sayur dari ibu balita stunting masih rendah. Kemudian 57% balita stunting belum mendapatkan pemberian makanan tambahan. Selain itu, hanya 4% balita *stunting* yang pernah didampingi oleh tenaga gizi, dan sisanya sebesar 96% belum pernah didampingi oleh tenaga gizi.
2. Berdasarkan hasil klasifikasi CART dengan ADASYN-N, faktor-faktor yang mengindikasikan terjadinya kasus *stunting* di Kota Surabaya adalah tinggi badan balita dibandingkan umur, ketersediaan jamban, dan tidak adanya anggota keluarga yang merokok di dalam rumah. Sedangkan berdasarkan hasil klasifikasi CART dengan ADASYN-kNN, faktor-faktor yang mengindikasikan terjadinya kasus *stunting* adalah tinggi badan dibandingkan umur, riwayat penyakit ibu hamil, pendidikan terakhir ibu balita, dan total pendapatan keluarga.
3. Klasifikasi CART dengan menggunakan penanganan data *imbalance* ADASYN-N memiliki nilai AUC terbaik yaitu sebesar 97,35%.

5.2 Saran

Saran yang dapat diberikan pada Dinas Kesehatan ataupun pihak lain yang bersangkutan berdasarkan hasil analisis adalah agar lebih memberikan edukasi kepada para ibu terkait kesadaran dalam pemenuhan gizi, seperti kesadaran untuk memakan buah dan sayur. Selain itu perlu dilakukan peningkatan

pemerataan pemberian makanan tambahan pada balita yang memiliki gizi kurang, peningkatan tenaga gizi pendamping bagi balita yang terindikasi mengalami gizi buruk juga perlu dilakukan. Penelitian lebih lanjut terkait dampak kebiasaan merokok anggota keluarga dapat mempengaruhi terjadinya *stunting* juga perlu dilakukan. Kemudian peningkatan sarana sanitasi juga harus diperhatikan terutama ketersediaan jamban sehat yang memenuhi syarat bagi setiap rumah tangga.

DAFTAR PUSTAKA

DJKM Kemenkes RI,. (2019, February 8). (Kementrian Kesehatan Republik Indonesia) Retrieved May 7, 2020, from Direktorat Jenderal Kesehatan Masyarakat: www.kesmas.kemkes.go.id/portal/konten/~rilis-berita/020819-bulan-vitamin-a

Andewi. (2011, February 14). *CFC Penatalaksanaan Gizi Buruk di Masyarakat*. (Kementrian Kesehatan Republik Indonesia) Retrieved May 7, 2020, from Direktorat Jenderal Kesehatan Masyarakat: www.kesmas.kemkes.go.id/portal/konten/~rilis-berita/021415-cfc-penatalaksanaan-gizi-buruk-di-masyarakat

Aritonang, N. E., Rusgiyono, A., & Rahmawati, R. (2016). Klasifikasi Status Kerja pada Angkatan Kerja Kota Semarang Tahun 2014 menggunakan Metode CHAID dan CART. *Jurnal Gaussian*, 183-192.

Astari, L. (2006). *Faktor-Faktor yang Berpengaruh terhadap Kejadian Stunting Anak Usia 6-12 Bulan di Kabupaten Bogor*. *Tesis*. Bogor: Sekolah Pascasarjana Institut Pertanian Bogor.

Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation Measure for Models Assessment over Imbalanced Data Sets. *Journal of Information Engineering and Aplications*.

Bekti, R. D., Pratiwi, N., Jatipaningrum, M. T., & Auliana, D. (2017). Analisis Pengaruh Lokasi dan Karakteristik Konsumen dalam Memilih Minimarket dengan Metode Regresi Logistik dan CART. *Media Statistika*, 119-130.

Breiman L., F. J. (1993). *Classification And Regression Trees*. New York: Chapman And Hall .

Choi, M. (2010). *A Selective Sampling Method for Imbalanced Data Learning on Support Vector Machines.* Iowa: Iowa State University.

Han, & Kamber. (2006). *Data Mining : Concepts and Techniques Second Edition.* San Francisco: Morgan Kaufmann.

He, H., Bai, Y., Garcia, E., & Li, S. (2008). ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learnin. *International Joint Conference on Neural Networks*.

Humas RSUP Dr. Sardjito Yogyakarta. (2019, July 22). *Kenali Penyebab Stunting Anak*. (RSUP Dr. Sardjito) Retrieved January 19, 2020, from RSUP Dr. Sardjito: https://sardjito.co.id/2019/07/22/kenali-penyebab-stunting-anak/

Kemenkes RI. (2018). *Petunjuk Teknis Pemberian Makanan Tambahan (Balita-Anak Sekolah-Ibu Hamil)*. Jakarta: Kementrian Kesehatan Republik Indonesia.

Kemenkes RI. (2018). Situasi Balita Pendek (Stunting) di Indonesia. *Data dan Informasi kesehatan*, 2.

Kemenkes RI, P. D. (2014). *Situasi dan Analisis Imunisasi*. Jakarta Selatan: Pusat Data dan Informasi Kemenkes RI.

Kementrian Kesehatan R. (2010). *Keputusan Menteri Kesehatan Republik Indonesia Nomor 1995/Menkes/SK/XII/2010 Tentang Standar Antropometri Penilaian Status Gizi Anak*. Jakarta: Direktorat Jenderal Bina Gizi dan Kesehatan Ibu dan Anak.

Kementrian Kesehatan RI. (2010). *Patent No. No. 1995/MENKES/SK/XII/2010*. Indonesia.

Kuntari, T., Jamil, N. A., & Sunarto, K. O. (2013). Faktor Risiko Malnutrisi pada Balita. *7*(12).

Kurniawati, Y. E. (2014). Adaptive Synthetic–Nominal (ADASYN–N) and Adaptive Synthetic–KNN (ADASYNKNN) for Multiclass Imbalance Learning on

Laboratory Test Data . *4th International Conference on Science and Technology (ICST)*, 1.

Larose, D. (2005). *Discovering Knowledge in Data : An Introduction to Data Mining.* New York: John Willey and Sons.

Lestari, W., Hartati, S., Mayasari, D., & Sakiyanto. (2018). Faktor yang Berhubungan dengan Kejadian Stunting pada Anak Sekolah Dasar Negeri 014610 Sei Renggas Kecamatan Kisaran Barat Kabupaten Asahan. *Jurnal Dunia Gizi*, 59-64.

Lewis, R. J. (2000). An Introduction to Classification and Regression Tree (CART) Analysis. *Presented at the 2000 Anual Meeting of Society For Academy Emergency Medicine in San Fransisco, California*.

Maryuni, A. (2012). *Inisiasi Menyusui Dini, ASI Eksklusif dan Manajemen Laktasi.* Jakarta: Trans Info Media.

Mustika, D. A. (2020, January 2). *Stunting dan Pusaran Kemiskinan*. Retrieved January 19, 2020, from SUARA.com: https://www.suara.com/yoursay/2020/01/02/094430/stunting-dan-pusaran-kemiskinan

P2PTM Kemenkes RI. (2018, Juny 25). *Konsumsi Rokok Akibatkan Anak Stunting*. (Kementrian Kesehatan RI) Retrieved May 6, 2020, from Direktorat P2PTM Kementerian Kesehatan RI: p2ptm.kemkes.go.id/kegiatan-p2ptm/pusat-/konsumsi-rokok-akibatkan-anak-stunting

Prawirohardjo, S. (2006). *Buku Acuan Nasional Pelayanan Kesehatan Maternal dan Neonatal.* Jakarta: Yayasan Bina Pustaka.

Rahayu, S., Aji, T. B., & Setiawan, N. A. (2017). *Penghitungan k-NN pada Adaptive Synthetic-Nominal (ADASYN-N) dan Adaptive Synthetic-kNN (ADASYN-kNN) untuk Data*

NominalMulti Kategori. Yogyakarta: Universitas Gadjah Mada.

Roesli, U. (2008). *Inisiasi Menyusu Dini Plus ASI Eksklusif* . Jakarta: Pustaka Bunda.

Satriani S, E. (2018). *Menyusui sebagai Dasar Kehidupan.* Jakarta Selatan: Pusat Data dan Informasi Kementrian Kesehatan RI.

Setiawan, E., Machmud, R., & Masrul. (2018). Faktor-Faktor yang Berhubungan dengan Kejadian Stunting pada Anak Usia 24-59 Bulan di Wilayah Kerja Puskesmas Andalas Kecamatan Padang Timur Kota Padang Tahun 2018 . *Jurnal Kesehatan Andalas*, 2.

Sudarsih, S., & Wijayanti. (2013). Hubungan Antara Pendapatan Keluarga Dengan Status Gizi Balita Usia 36-60 Bulan Di Wilayah Kerja Puskesmas Gondang Mojokerto. *MEDICA MAJAPAHIT*, 5(2).

Trihono, A. T. (2015). *Pendek (stunting) di Indonesia, Masalah dan Solusinya.* Jakarta: Lembaga Penerbit Balitbangkes.

Umakaapa, M. (2020, January 14). *detikNews*. (detiknetwork) Retrieved January 19, 2020, from detikcom: https://news.detik.com/kolom/d-4858547/strategi-2020-melawan-stunting

Waluyo, A., Mukid, M. A., & Wuryandari, T. (2014). Perbandingan Klasifikasi Nasabah Kredit Menggunakan Regresi Logistik Biner dan Cart (Classification And Regression Trees). *Media Statistika*, 95-104.

Wang, S., & Yao, X. (2013). Using Class Imbalance Learning for Software Defect Prediction. *IEEE Transactions on Reliability*, 62, 438.

Waryana. (2010). *Gizi Reproduksi.* Yogyakarta: Pustaka Rihama.

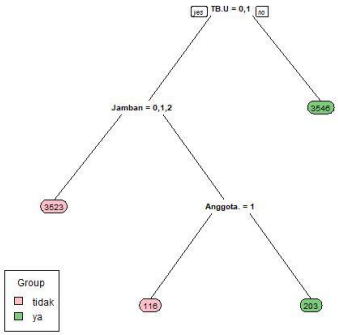
LAMPIRAN

Lampiran 1. Data Penelitian

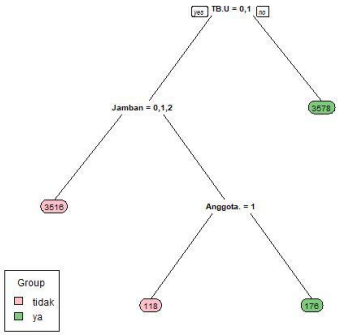
No	X1	X2	X3	X4	X5	…	X24	X25	X26	Y
1	5	0	1	2	1		0	1	0	0
2	5	1	0	2	1		1	1	1	0
3	0	1	1	1	0		1	1	1	0
4	0	1	0	2	2		0	0	1	0
5	0	1	1	2	1		1	1	1	0
6	0	1	0	1	1		0	1	1	0
7	0	1	1	1	0		1	1	0	0
8	0	1	1	2	1		1	0	1	0
9	0	1	0	1	0		1	1	1	0
10	0	0	1	2	1		0	0	1	0
11	0	0	1	2	1		1	0	0	0
12	0	1	0	2	1		1	1	1	0
13	0	1	1	2	1		0	1	1	0
14	1	0	0	2	2		0	1	1	0
15	0	0	1	0	0		1	0	1	0
16	0	1	0	2	1		1	0	1	0
17	0	1	1	2	1		0	1	0	0
18	5	0	0	2	1		1	1	1	0
19	5	0	1	2	2		0	0	0	0
4750	0	0	0	0	0		1	1	1	0
4751	0	0	1	1	0		1	0	0	0
4752	0	1	1	2	1		1	1	1	0
4753	5	0	0	1	1		0	0	1	0
4754	0	0	1	2	1	…	0	0	1	0

Lampiran 2. Pohon Klasifikasi CART dengan ADASYN-N

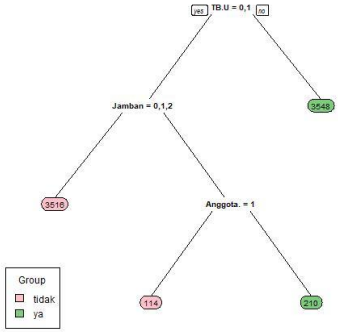
(1)



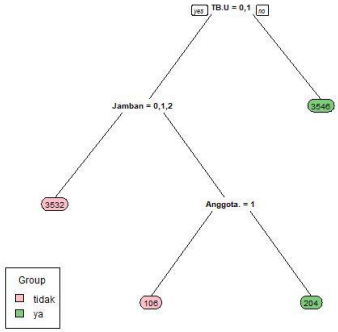
(2)



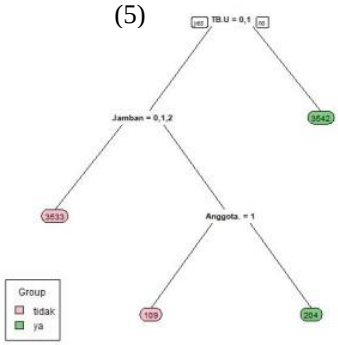
(3)



(4)

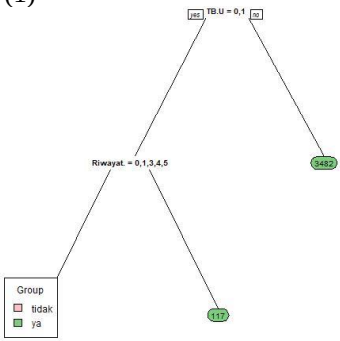


(5)

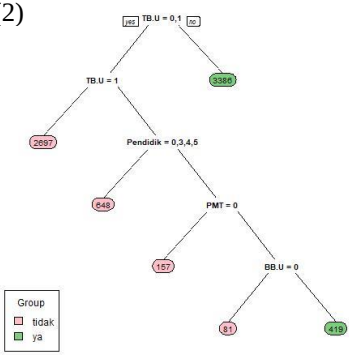


Lampiran 3. Pohon Klasifikasi CART dengan ADASYN-kNN

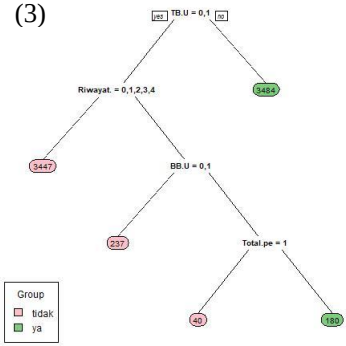
(1)



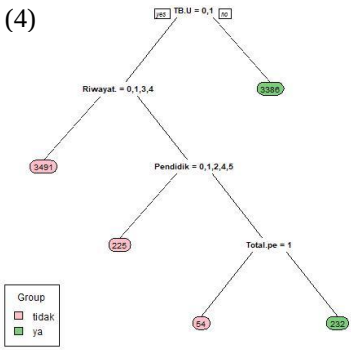
(2)



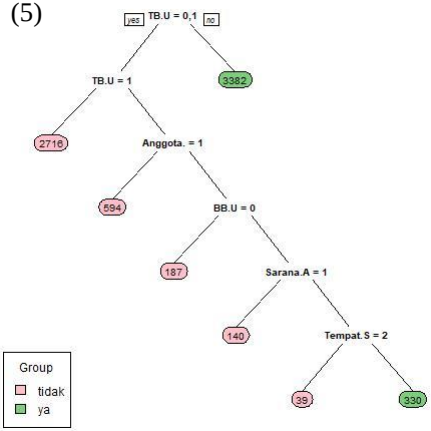
(3)



(4)



(5)


yes
TB.U = 0,1
no
TB.U = 1
3382
2716
Anggota. = 1
594
BB.U = 0
187
Sarana.A = 1
140
Tempat.S = 2
39
330
Group
tidak
ya

Lampiran 4. *Syntax* R CART dengan ADASYN-N

```
data <- read.csv("E:/Skripsi27.csv", sep=";")
setwd("E:/ITS")
set.seed(12345)
library(UBL)
library(caret)
library(rpart)
library(rpart.plot)
library(dplyr)

#factorized
for (i in 1:ncol(data)) {
  data[,i] <- as.factor(data[,i])
}

class.freq <- function (dat, tgt)
{
  names <- sort(unique(dat[, tgt]))
  li <- list(names, sapply(names, function(x)
length(which(dat[, tgt] == x))))
  li
}
train.index <- createDataPartition(data$Stunting, times =
1, p = 0.8, list = F)
training <- data[train.index,]
testing <- data[-train.index,]

  li <- class.freq(training, "Stunting")
  tgt <- which(training$Stunting == li[[1]][2])
  ms <- length(tgt)
  ml <- nrow(training) - ms
  d <- ms/ml
  q <- ncol(training)
  G <- (ml-ms)

  #menghitung jarak amatan kelas minor seluruh data
  kNNs <- neighbours("Stunting", training, dist="HVDM",
k=10)
```

```
er <- array(0,c(length(tgt),1))

  for (i in 1:length(tgt)) {
    r <- 0
    for (j in 1:10) {
      r <- r+length(which(training[kNNs[tgt[i],j],27] ==
li[[1]][1]))
    }
    er[i] <- r/10
  }

  newr <- array(0,c(length(tgt),1))
  g <- array(0,c(length(tgt),1))
  for (k in 1:length(tgt)) {
    newr[k] <- er[k]/sum(er[,1])
    g[k] <- round(newr[k]*G)
  }

  datamin <- array(0,c(length(tgt),q))
  datamin <- training[training$Stunting == li[[1]][2],]

  datamin <- cbind(datamin,g)
  datamin$g <- as.numeric(datamin$g)

  #replicate
  datamin <- datamin[rep(seq(nrow(datamin)),
datamin$g),]
  datamin <- datamin[ , !(names(datamin) %in% "g")]

  databaru <- rbind(training,datamin)
  databaru

for (t in 1:1) {
    jpeg(file = paste("Adasyn-N setseed12345",t,'.jpeg',
sep = ''))

# grow tree
fit <- rpart(Stunting ~ .,method="class", data=databaru)
```

```
#pruned tree
bestcp <-fit$cptable[which.min(fit$cptable
                                [,"xerror"]),"CP"]
tree.pruned <- prune(fit, cp = bestcp)

# plot tree
plot(tree.pruned)
text(tree.pruned, cex = 0.8, use.n = TRUE, xpd = TRUE)
prp(tree.pruned, faclen = 0, cex = 0.8, extra = 1)
tot_count <- function(x, labs, digits, varlen)
{
  paste(labs, "\n\nn =", x$frame$n)
}

prp(tree.pruned, faclen = 0, cex = 0.8,
node.fun=tot_count)
only_count <- function(x, labs, digits, varlen)
{
  paste(x$frame$n)
}

boxcols <- c("pink", "palegreen3")[tree.pruned$frame$yval]

par(xpd=TRUE)
prp(tree.pruned, faclen = 0, cex = 0.8,
node.fun=only_count, box.col = boxcols)
legend("bottomleft", legend = c("tidak","ya"), fill =
c("pink", "palegreen3"),
       title = "Group")

dev.off()

#predict test data
cartpredict <- tree.pruned %>% predict(testing,
type="class")
```

```
# confusion matrix (training data) cart
conf.matrixt1 <- table(databaru$Stunting,
predict(tree.pruned,type="class"))
rownames(conf.matrixt1) <- paste("Actual",
rownames(conf.matrixt1), sep = ":")
colnames(conf.matrixt1) <- paste("Pred",
colnames(conf.matrixt1), sep = ":")
print(conf.matrixt1)
# confusion matrix (testing data) cart
conf.matrix <- table(testing$Stunting, cartpredict)
rownames(conf.matrix) <- paste("Actual",
rownames(conf.matrix), sep = ":")
colnames(conf.matrix) <- paste("Pred",
colnames(conf.matrix), sep = ":")
print(conf.matrix)

auc <- array(0,c(1,1))
auctr <- array(0,c(1,1))

#auc training
auctr <-
((conf.matrixt1[2,2]/(conf.matrixt1[2,1]+conf.matrixt1[2
,2]))+(conf.matrixt1[1,1]/(conf.matrixt1[1,2]+conf.matri
xt1[1,1])))/2

#auc testing
auc <-
((conf.matrix[2,2]/(conf.matrix[2,1]+conf.matrix[2,2]))+
(conf.matrix[1,1]/(conf.matrix[1,2]+conf.matrix[1,1])))/
2
  }
```

Lampiran 5. *Syntax* R CART dengan ADASYN-kNN

```
data <- read.csv("E:/Skripsi27.csv", sep=";")
setwd("E:/ITS")
set.seed(12345)
library(UBL)
library(caret)
library(rpart)
library(rpart.plot)
library(dplyr)

#factorized
for (i in 1:ncol(data)) {
  data[,i] <- as.factor(data[,i])
}

auc <- array(0,c(1,1))
auctr <- array(0,c(1,1))

class.freq <- function (dat, tgt)
{
  names <- sort(unique(dat[, tgt]))
  li <- list(names, sapply(names, function(x)
length(which(dat[, tgt] == x))))
  li
}

listdat <- list()
train.index <- createDataPartition(data$Stunting,
times = 1, p = 0.8, list = F)
training <- data[train.index,]
testing <- data[-train.index,]

listdat[[1]] <- AdasynClassif(Stunting ~ ., training,
baseClass = NULL, beta = 1, k=10, dist = "HVDM")

for (t in 1:1) {
  jpeg(file = paste("adasyn-knn
setseed12345",t,'.jpeg', sep = ''))
```

```
# grow tree
fit <- rpart(Stunting ~ .,method="class", data=databaru)
#pruned tree
bestcp <-fit$cptable[which.min(fit$cptable
                              [,"xerror"]),"CP"]
tree.pruned <- prune(fit, cp = bestcp)

# plot tree
plot(tree.pruned)
text(tree.pruned, cex = 0.8, use.n = TRUE, xpd = TRUE)
prp(tree.pruned, faclen = 0, cex = 0.8, extra = 1)
tot_count <- function(x, labs, digits, varlen)
{
  paste(labs, "\n\nn =", x$frame$n)
}

prp(tree.pruned, faclen = 0, cex = 0.8,
node.fun=tot_count)
only_count <- function(x, labs, digits, varlen)
{
  paste(x$frame$n)
}

boxcols <- c("pink",
"palegreen3")[tree.pruned$frame$yval]

par(xpd=TRUE)
prp(tree.pruned, faclen = 0, cex = 0.8,
node.fun=only_count, box.col = boxcols)
legend("bottomleft", legend = c("tidak","ya"), fill =
c("pink", "palegreen3"),
       title = "Group")

dev.off()

#predict test data
cartpredict <- tree.pruned %>% predict(testing,
type="class")
```

```
# confusion matrix (training data) cart
conf.matrixt1 <- table(databaru$Stunting,
predict(tree.pruned,type="class"))
rownames(conf.matrixt1) <- paste("Actual",
rownames(conf.matrixt1), sep = ":")
colnames(conf.matrixt1) <- paste("Pred",
colnames(conf.matrixt1), sep = ":")
print(conf.matrixt1)
# confusion matrix (testing data) cart
conf.matrix <- table(testing$Stunting, cartpredict)
rownames(conf.matrix) <- paste("Actual",
rownames(conf.matrix), sep = ":")
colnames(conf.matrix) <- paste("Pred",
colnames(conf.matrix), sep = ":")
print(conf.matrix)

auc <- array(0,c(1,1))
auctr <- array(0,c(1,1))

#auc training
auctr <-
((conf.matrixt1[2,2]/(conf.matrixt1[2,1]+conf.matrixt1[2
,2]))+(conf.matrixt1[1,1]/(conf.matrixt1[1,2]+conf.matri
xt1[1,1])))/2

#auc testing
auc <-
((conf.matrix[2,2]/(conf.matrix[2,1]+conf.matrix[2,2]))+
(conf.matrix[1,1]/(conf.matrix[1,2]+conf.matrix[1,1])))/
2
  }
```

Lampiran 6. Surat Keterangan

PEMERINTAH KOTA SURABAYA
DINAS KESEHATAN
Jalan Jemursari No. 197 Surabaya 60243
Telp. (031) 8439473, 8439372, 8473729 Fax. (031) 8483393

SURAT IJIN
SURVEY / PENELITIAN
Nomor : 072 / 1939 / 436.7.2 / 2020

Dari	:	Sekretaris Kepala Badan Kesatuan Bangsa, Politik dan Perlindungan Masyarakat
Nomor	:	070/2408/436.8.5/2020
Tanggal	:	14 Februari 2020
Hal	:	Keterangan Selesai Pegambilan Data

Dengan ini menyatakan tidak keberatan dilakukan survey / penelitian oleh :

Nama	:	**Winantri Nurbuwah Insyroh**
NIM	:	06211640000034
Pekerjaan	:	Mahasiswa Statistika Fakultas Sains dan Analitika Data ITS Kota Surabaya
Alamat	:	Dsn. Wonosari Tengah Kec. Turur Kab. Pasuruan
Tujuan Penelitian	:	Menyusun Proposal
Tema Penelitian	:	Kalsifikasi Faktor-Faktor yang Menyebabkan balita Stunting di Kota Surabaya Menggunakan Metode Classification and Regreession Trees (CART)
Lamanya Penelitian	:	Bulan Februari s/d Bulan Mei Tahun 2020
Daerah / tempat Penelitian	:	Dinas Kesehatan Kota Surabaya

Dengan syarat – syarat / ketentuan sebagai berikut :
1. Yang bersangkutan harus mentaati ketentuan-ketentuan/ peraturan yang berlaku dimana dilakukannya kegiatan survey/penelitian.
2. Dilarang menggunakan kuesioner diluar design yang telah ditentukan.
3. Yang bersangkutan sebelum dan sesudah melakukan survey/penelitian harap melaporkan pelaksanaan dan hasilnya kepada Dinas Kesehatan Kota Surabaya.
4. Surat ijin ini akan dicabut/tidak berlaku apabila yang bersangkutan tidak memenuhi syarat-syarat serta ketentuan seperti diatas.

Sehubungan dengan hal tersebut diharapkan kepada Saudara Kepala Puskesmas untuk memberikan bantuan, pengarahan dan bimbingan sepenuhnya.
Demikian atas perhatian Saudara disampaikan terima kasih.

Surabaya, 19 Mei 2020
a.n. KEPALA DINAS
Sekretaris,

Nanik Sukristina S.KM. M.Kes
Pembina Tk. I
NIP. 19700117199403 2008

http://dinkes.surabaya.go.id, Email : dinkes.surabaya@gmail.com

BIODATA PENULIS

Penulis dilahirkan di Pasuruan, 4 Agustus 1997 dengan nama lengkap Winantri Nurbuwah Insyroh, dan biasa dipanggil Inan. Penulis menempuh pendidikan formal di SD Negeri 1 Nongkojajar, SMP Negeri 1 Tutur, dan SMA Negeri 1 Lawang. Kemudian penulis masuk sebagai mahasiswa Departemen Statistika ITS pada tahun 2016. Selama masa perkuliahan, penulis aktif di Departemen Sosial Masyarakat Himpunan Mahasiswa Statistika ITS (HIMASTA-ITS) sebagai staff pada periode 2017/2018, dan aktif menjadi Ketua Biro Pengabdian Masyarakat, Lingkungan Kesehatan, dan Kajian Strategis di Departemen Sosial Masyarakat HIMASTA-ITS periode 2018/2019. Selain itu, penulis juga aktif dalam kegiatan kepanitiaan seperti menjadi Koordinator Konsumsi dan Kesehatan Pekan Raya Statistika 2017 (PRS 2017), panitia di Indonesia Statistics Conference and Olympiad 2017 (ISCO 2017), dan beberapa event lainnya. Bagi pembaca yang ingin berdiskusi, memberikan saran, dan kritik mengenai Tugas Akhir ini dapat disampaikan melalui email winaninsyrh@gmail.com atau ke nomor telepon 082230919878.