

**PENGLASIFIKASIAN MENGGUNAKAN *CHI-SQUARE AUTOMATIC
INTERACTION DETECTION*
DENGAN PENDEKATAN *SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE***

(Studi Kasus Bayi Berat Lahir Rendah (BBLR) di Jawa Barat)

SKRIPSI

Diajukan untuk Memenuhi Sebagian dari Syarat Memperoleh

Gelar Sarjana Matematika



Oleh:

Ivo Oktaviano

1703862

**PROGRAM STUDI MATEMATIKA
DEPARTEMEN PENDIDIKAN MATEMATIKA
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA**

2021

Ivo Oktaviano, 2021

***PENGLASIFIKASIAN MENGGUNAKAN *CHI-SQUARE AUTOMATIC INTERACTION DETECTION*
DENGAN PENDEKATAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (Studi Kasus Bayi
Berat Lahir Rendah (BBLR) di Jawa Barat)***

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

**PENGLASIFIKASIAN MENGGUNAKAN *CHI-SQUARE AUTOMATIC
INTERACTION DETECTION*
DENGAN PENDEKATAN *SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE*
(Studi Kasus Bayi Berat Lahir Rendah (BBLR) di Jawa Barat)**

Oleh
Ivo Oktaviano

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Matematika pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

© Ivo Oktaviano 2021
Universitas Pendidikan Indonesia
April 2021

Hak Cipta dilindungi undang-undang.

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak ulang, difotokopi, atau lainnya tanpa izin dari peneliti

LEMBAR PENGESAHAN

IVO OKTAVIANO

PENGLASIFIKASIAN MENGGUNAKAN *CHI-SQUARE AUTOMATIC INTERACTION DETECTION* DENGAN PENDEKETAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (Studi Kasus Bayi Berat Lahir Rendah (BBLR) di Jawa Barat)

Disetujui dan disahkan oleh pembimbing:

Pembimbing I



Drs. Nar Herrhyanto, M.Pd.

NIP. 196106181987031001

Pembimbing II




Fitriani Agustina, S.Si., M.Si.

NIP. 198108142005012001

Mengetahui,

Ketua Departemen Pendidikan Matematika



Dr. H. Dadang Juandi, M.Si.

NIP. 196401171992021001

i

Ivo Oktaviano, 2021

PENGLASIFIKASIAN MENGGUNAKAN *CHI-SQUARE AUTOMATIC INTERACTION DETECTION* DENGAN PENDEKETAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (Studi Kasus Bayi Berat Lahir Rendah (BBLR) di Jawa Barat)

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

**PENGLASIFIKASIAN MENGGUNAKAN *CHI-SQUARE AUTOMATIC
INTERACTION DETECTION*
DENGAN PENDEKATAN *SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE***

(Studi Kasus Bayi Berat Lahir Rendah (BBLR) di Jawa Barat)

ABSTRAK

Metode *Chi-Square Automatic Interaction Detection* (CHAID) merupakan suatu teknik iteratif yang menguji satu persatu variabel independen yang digunakan dalam klasifikasi, dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya serta *output* dari analisis CHAID berupa diagram pohon klasifikasi yang menggambarkan pembentukan segmen. Keseimbangan data perlu diperhatikan dalam penerapan metode CHAID, karena ketidakseimbangan data (*Imbalanced Data*) mengakibatkan kesalahan dalam klasifikasi. Solusi untuk mengatasi ketidakseimbangan data adalah dengan menerapkan metode *Synthetic Minority Oversampling Technique* (SMOTE). Dalam penelitian ini, metode CHAID dengan pendekatan SMOTE diterapkan pada kasus Bayi Berat Lahir Rendah (BBLR) di Jawa Barat yang belum mencapai target SDGs, untuk mengetahui variabel-variabel yang mencirikan BBLR di Jawa Barat. Tujuan dari penelitian ini adalah untuk mengetahui variabel-variabel yang mencirikan BBLR di Jawa Barat dengan membandingkan antara metode CHAID dan metode CHAID dengan pendekatan SMOTE. Hasil dari penelitian ini menunjukkan bahwa metode CHAID dengan pendekatan SMOTE lebih baik daripada metode CHAID tanpa pendekatan SMOTE untuk mencirikan kejadian BBLR di Jawa Barat.

Kata Kunci: BBLR, CHAID, *Imbalanced Data*, SMOTE.

**CLASSIFICATION USING CHI-SQUARE AUTOMATIC
INTERACTION DETECTION**
WITH SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE
APPROACH
(Case Study of Low Birth Weight Infants in West Java)

ABSTRACT

Chi-Square Automatic Interaction Detection method is an iterative technique which test one by one independent variables used in the classification, and arrrage them based on the level of chi-square statistical significance of dependent variable. Output of the CHAID analysis is a classification tree diagram which describes segment formation. The balance of the data must be considered in the application of the CHAID method, because imbalanced data result in error in classification. The solution to solve imbalanced data is to apply the Synthetic Minority Oversampling Technique (SMOTE) method. In this study, CHAID method with SMOTE approach was applied to cases of Low Birth Weight (LBW) Babies in West Java who have not reached the SDGs target, to know the variables that characterize LBW in West Java. The purpose of this study is to know the variables that characterize LBW in West Java by comparing the CHAID method and CHAID method with SMOTE approach. The results of this study indicate that the CHAID method with SMOTE approach is better than CHAID method without SMOTE approach for characterizing the incidance of LBW in West Java.

Keywords: *Low Birth Weight (LBW) Babies, CHAID, Imbalanced Data, SMOTE.*

DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN KEASLIAN ISI SKRIPSI	ii
KATA PENGANTAR	iii
UCAPAN TERIMA KASIH.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN.....	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.5.1 Manfaat Teoritis	5
1.5.2 Manfaat Praktis	5
BAB II KAJIAN TEORI.....	6
2.1 Berat Badan Lahir	6
2.2 Tabel Kontingensi (<i>Contingency Table</i>)	11
2.3 Data Tidak Seimbang (<i>Imbalanced Data</i>).....	12
2.4 Strategi Sampling	12
2.5 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	13
2.6 Data Mining.....	15
2.7 <i>Chi-Squared Automatic Interaction Detection (CHAID)</i>	16
2.8 Algoritma <i>Chi-Squared Automatic Interaction Detection (CHAID)</i>	17
2.8.1 <i>Merging</i> (Penggabungan).....	17
2.8.2 <i>Splitting</i> (Pemisahan)	19

2.8.3 <i>Stopping</i> (Penghentian)	19
2.9 Koreksi Bonferroni (<i>Bonferroni Correction</i>)	20
2.10 Diagram Pohon Klasifikasi CHAID (<i>CHAID Classification Tree</i>).....	21
2.11 Ukuran Ketepatan Klasifikasi.....	22
2.12 Hasil Penelitian Sebelumnya.....	24
BAB III METODOLOGI PENELITIAN.....	26
3.1 Metode Penelitian.....	26
3.2 Metode <i>Chi-Square Automatic Interaction Detection</i> dan Metode <i>Synthetic Minority Oversampling Technique</i>	27
BAB IV HASIL DAN PEMBAHASAN	30
4.1 Data	30
4.2 Gambaran Umum Karakteristik BBLR di Jawa Barat Periode 2012-2017	31
4.3 Analisis CHAID pada BBLR	36
4.4 Analisis CHAID pada BBLR dengan Pendekatan SMOTE.....	40
4.5 Perbandingan CHAID dan CHAID dengan Pendekatan SMOTE	45
BAB V KESIMPULAN DAN SARAN.....	47
5.1 Kesimpulan.....	47
5.2 Saran.....	48
DAFTAR ISI	49
LAMPIRAN.....	53

DAFTAR TABEL

Tabel 2.1 Tabel Kontingensi I x J	11
Tabel 2.2 Tabel Probabilitas Kontingensi I x J	12
Tabel 2.3 Confusion Matrix	23
Tabel 2.4 Hasil Penelitian Sebelumnya	24
Tabel 4.1 Persentase Usia Ibu Terhadap Kejadian BBLR.....	31
Tabel 4.2 Persentase Pendidikan Ibu Terhadap Kejadian BBLR	32
Tabel 4.3 Persentase Tingkat Ekonomi Ibu Terhadap Kejadian BBLR	33
Tabel 4.4 Persentase Paritas Ibu Terhadap Kejadian BBLR.....	33
Tabel 4.5 Persentase Tempat Tinggal Ibu Terhadap Kejadian BBLR.....	34
Tabel 4.6 Persentase Status Bekerja Ibu Terhadap Kejadian BBLR	35
Tabel 4.7 Persentase Kebiasaan Merokok Ibu Terhadap Kejadian BBLR	35
Tabel 4. 8 Hasil Segmentasi CHAID pada Kejadian BBLR.....	38
Tabel 4.9 Nilai Indeks CHAID	39
Tabel 4.10 Hasil Klasifikasi Pada Kejadian BBLR	39
Tabel 4.11 Hasil Segmentasi CHAID dengan Pendekatan SMOTE pada Kejadian BBLR	43
Tabel 4.12 Nilai Indeks CHAID dengan Pendekatan SMOTE.....	44
Tabel 4.13 Hasil Klasifikasi Kejadian BBLR dengan Pendekatan SMOTE	45
Tabel 4.14 Ketepatan Klasifikasi	45

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Algoritma SMOTE.....	14
Gambar 2.2 Diagram Pohon CHAID.....	21
Gambar 3.1 Diagram Alir Penelitian	26
Gambar 4.1 Diagram Pohon CHAID Pada Kejadian BBLR	37
Gambar 4.2 Diagram Pohon CHAID dengan Pendekatan SMOTE Pada Kejadian BBLR	41

DAFTAR LAMPIRAN

Lampiran 1. Definisi Operasional	53
Lampiran 2. Filter dan Cleaning Data	58
Lampiran 3. Data Kategorik (Responden Wanita Usia Subur 15-49 di Jawa Barat (SDKI 2017))	58
Lampiran 4. Nilai Indeks CHAID	77
Lampiran 5. Nilai Risk Estimate CHAID	77
Lampiran 6. Tabel Klasifikasi CHAID	77
Lampiran 7. Data Kategorik Hasil Metode SMOTE	78
Lampiran 8. Nilai Indeks CHAID dengan Pendekatan SMOTE	112
Lampiran 9. Nilai Risk Estimate CHAID dengan Pendekatan SMOTE.....	112
Lampiran 10. Tabel Klasifikasi CHAID dengan Pendekatan SMOTE.....	112

DAFTAR PUSTAKA

- A. C, Flores., R. I, Icoy., C. F, Pena., & K. D, Gorro.. (2018). An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set. *International Conference on Engineering Applied Sciences an Technplogy (ICEAST)*, 1-4.
- Ahsan, M., Gomes, R., & Denton, A. (2018). SMOTE Implementation on Phising Data to Enhance Cybersecurity . *International Conference on Electro/Information Technology (EIT)*, 531-536.
- Ariawan, I. (2006). Indeks sosio-ekonomi menggunakan Principal Componen Analysis (PCA). *Jurnal Kemas UI*. Diambil kembali dari <http://jurnalkemas.ui.ac.id/kesmas/article/view/317/316> [15 September 2020]
- Arifiyanti, A. A. (2020). Metode Penyeimbang Kelas Pada Klasifikasi Data Mining. *UPN Veteran*, 34.
- Arikunto, S. (2006). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rhineka Cipta.
- Badan Perencanaan Pembangunan Nasional. (2015). *Rencana Pembangunan Jangka Menengah Nasional*. Jakarta: BAPPENAS.
- Badan Perencanaan Pembangunan Nasional. (2015). *Rencana Pembangunan Jangka Menengah Nasional 2015-2019*. Jakarta: BAPPENAS.
- Badan Pusat Statistik, Badan Kependudukan dan Keluarga Berencana Nasional, Kementerian Kesehatan , & USAID. (2018). *Survei Demografi dan Kesehatan Indonesia 2017*. Jakarta: BPS. BKKBN, Kementerian Kesehatan, USAID.
- Barro, R., & Itasia, S. (2013). Penerapan SMOTE Terhadap Data Tidak Seimbang. *Journal Of Statistics*.

- Chawla, V. N., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*.
- Departemen Kesehatan RI. (2003). *Indikator Indonesia Sehat dan Penetapan Indikator Provinsi Sehat dan Kabupaten/Kota Sehat*. Jakarta: Departemen Kesehatan.
- Eherler, D., & Lehmann, T. (2001). Responder Profiling with CHAID and Dependency Analysis. *European Conference on Machine Learning, 12*, 49-58.
- Gallagher, C. A., Monroe, H. M., & Fish, J. L. (2000). An Iterative Approach to Classification Analysis. *Casualty Actuarial Society*, 237-280.
- Haryati, dkk. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5. *Jurnal Media Infotama*, Vol. 11. No. 2. ISSN: 1858-2680.
- Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Andi Offset.
- Himawan, A. (2006). *Hubungan antara Karakteristik Ibu dengan Status Gizi Balita*. Semarang: Universitas Negeri Semarang.
- Kasali, R. (1998). *Membidik Pasar Indonesia*. Jakarta: PT Gramedia Pustaka Utama.
- Kementerian Kesehatan RI. (2015). *Rencana Strategis Kementerian Kesehatan Tahun 2015-2019*. Jakarta: Kementerian Kesehatan RI.
- Kunto, Y. S., & Hasana, S. N. (2006). Analisis CHAID Sebagai Alat Bantu Statistika untuk Segmentasi Pasar (Studi Kasus pada Koperasi Syari'ah Al-Hidayah). *Jurnal Manajemen Pemasaran*, 88-98.
- Larose, D. T. (2005). *Discovering Knowledge In Data: An Introduction to Data mining*. New Jersey: John Willey&Sons. Inc.

- Manuaba. (2010). *Ilmu Kebidanan Penyakit Kandungan*. Jakarta: EGC.
- Nooraeni, R. (2015). *Data Mining dengan R Konsep Serta Implementasi*. Jakarta: InMedia.
- Permana, H. (2011). *Klasifikasi dengan Metode CHAID (Chi-Squared Automatic Interaction Detection) dan Penerapannya pada Klasifikasi Alumni FMIPA UNY*. Yogyakarta.
- Prawirohardjo. (2007). *Ilmu Kebidanan (edisi pertama cetakan ketujuh)*. Jakarta: Yayasan Bina Pustaka Sarwono Prawirohardjo.
- Proverawati, A., & Ismawati, C. (2010). *Berat Badan Lahir Rendah (BBLR)*. Yogyakarta: Nuha Medika.
- R. I, R., & R. M, R. (2014). Data Mining Approaches to Predict Final Grade by Overcoming Class Imbalance Problem. *International Conference on Computer and Information Technology*, 14-19.
- Saifuddin, A. (2002). *Buku Acuan Nasional Pelayanan Kesehatan Maternal dan Neonatal*. Jakarta: Yayasan Bina Pustaka Sarwono Prawirohardjo.
- Sirrajuddin. (2011). Pengaruh Paparan Asap Rokok terhadap Kejadian Berat Badan Lahir Bayi di Sulawesi Selatan. *Jurnal Pendidikan Bidan*. Diambil kembali dari <http://jurnalpendidikanbidan.com> [15 September 2020]
- Sistriani, C. (2008). *Faktor Maternal dan Kualitas Pelayanan Antenatal yang Berisiko terhadap Kejadian BBLR. [Tesis]*. Semarang: FKM Universitas Diponegoro.
- Soetjiningsih. (2000). *Tumbuh Kembang Anak*. Jakarta: EGC.
- Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical Data Analysis Using The SAS System (Second Edition)*. North Carolina: SAS Institute Inc.
- Supariasa. (2001). *Penilaian Status Gizi*. Jakarta: Buku Kedokteran EGC.

UU RI Nomor 36 Tahun 2009 Tentang Kesehatan. (t.thn.).

Walpole, R. E. (1995). *Pengantar Statistika*. Jakarta: Gramedia Pustaka Utama.

WHO. (t.thn.). *Maternal Mortality*. World Health Organization.

Witten, I. H., & Frank, E. (2005). *DATA MINING Practical Machine Learning Tools and Techniques*. California: Morgan Kaufman.

Yuliva, I. D. (2009). Hubungan Satus Pekerjaan Ibu dengan Berat Lahir Bayi di RSUP dr. M. Djamil Padang. *Jurnal Berita Kedokteran Masyarakat*.

