

**IMPLEMENTASI *LOGISTIC REGRESSION ENSEMBLE*  
(LORENS) DALAM PENGKLASIFIKASIAN ANGKA  
KEMATIAN BAYI DI INDONESIA**

**SKRIPSI**

**OLEH  
MUHAMAD SABIT MUNAWAR  
NIM 19610005**



**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**IMPLEMENTASI *LOGISTIC REGRESSION ENSEMBLE*  
(LORENS) DALAM PENGKLASIFIKASIAN ANGKA  
KEMATIAN BAYI DI INDONESIA**

**SKRIPSI**

**Diajukan Kepada  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
untuk Memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Matematika (S.Mat.)**

**Oleh  
Muhamad Sabit Munawar  
NIM 19610005**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**IMPLEMENTASI *LOGISTIC REGRESSION ENSEMBLE*  
(LORENS) DALAM PENGKLASIFIKASIAN ANGKA  
KEMATIAN BAYI DI INDONESIA**

**SKRIPSI**

**Oleh  
Muhamad Sabit Munawar  
NIM 19610005**

Telah Diperiksa dan Disetujui untuk Diuji  
Malang, 20 Juni 2023

Dosen Pembimbing I



Ria Dhea Layla N. K., M.Si.  
NIDT. 19900709 20180201 2 228

Dosen Pembimbing II



Mohammad Nafie Jauhari, M.Si.  
NIDT. 19870218 20160801 1 056

Mengetahui,  
Ketua Program Studi Matematika



  
Dr. Efly Susanti, M.Sc.  
NIP. 19741129 200012 2 005

**IMPLEMENTASI *LOGISTIC REGRESSION ENSEMBLE*  
(LORENS) DALAM PENGKLASIFIKASIAN ANGKA  
KEMATIAN BAYI DI INDONESIA**

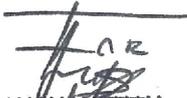
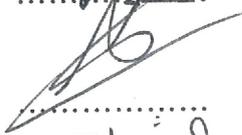
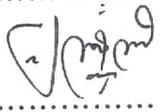
**SKRIPSI**

Oleh  
**Muhamad Sabit Munawar**  
NIM 19610005

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima sebagai salah satu Persyaratan  
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal, 22 Juni 2023

Ketua Penguji : Fachrur Rozi, M.Si.  
Anggota Penguji 1 : Angga Dwi Mulyanto, M.Si.  
Anggota Penguji 2 : Ria Dhea Layla N. K., M.Si.  
Anggota Penguji 3 : Mohammad Nafie Jauhari, M. Si.

  
.....  
  
.....  
  
.....  
  
.....

Mengetahui,  
Ketua Program Studi Matematika



Dr. Elfy Susanti, M.Sc.  
NIP. 19741129 200012 2 005

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhamad Sabit Munawar

NIM : 19610005

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Implementasi *Logistic Regression Ensemble* (LORENS)

dalam Pengklasifikasian Angka Kematian Bayi di Indonesia

Menyatakan dengan sebenar-benarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 22 Juni 2023

Yang membuat pernyataan,



Muhamad Sabit Munawar  
NIM. 19610005

## **MOTO**

“Kebodohan itu merusak, tetapi merasa dirinya paling pintar lebih merusak”

~Gus Baha~

## **PERSEMBAHAN**

*Bismillahirrahmanirrahim, alhamdulillahil'alamin*, dengan mengucap syukur kepada Allah Swt. skripsi ini peneliti persembahkan untuk:

Ayahanda tercinta Sudirman, ibunda tercinta Umiyah, pakde tercinta Nasimun Zainudin, bude tercinta Siti Roibah, semua saudara/i tersayang, Bu Ria Dhea Layla N. K. selaku dosen pembimbing, KKM DR ~ Vacation Desa Tlogosari yang senantiasa memberi dukungan moril kepada peneliti baik moril maupun materiil bagi peneliti.

## KATA PENGANTAR

Puji syukur kehadirat Allah Swt. atas segala limpahan rahmat dan karunia-Nya sehingga peneliti dapat menyelesaikan proposal skripsi dengan baik. Sholawat serta salam senantiasa tercurahkan kepada junjungan besar Nabi Muhamad saw. yang telah membawa dari jalan gelap gulita, yakni zaman jahiliah menuju jalan yang terang benderang, yakni *ad-dinul* Islam (agama Islam).

Proses penyusunan skripsi ini dapat selesai karena bimbingan dan arahan dari berbagai pihak. Untuk itu, ucapan terimakasih yang sebesar-besarnya dan penghargaan yang setinggi-tingginya penulis sampaikan kepada:

1. Prof. Dr. H. M. Zainuddin, MA. selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini, M.Si. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Elly Susanti, M.Sc. selaku Ketua Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Ria Dhea Layla N. K., M.Si. selaku dosen pembimbing 1 yang memberikan pengetahuan, arahan, nasihat, dan motivasi kepada penulis.
5. Mohammad Nafie Jauhari, M.Si. selaku dosen pembimbing 2 yang telah memberikan bimbingan, nasihat, dan ilmu kepada penulis.
6. Fachrur Rozi, M.Si. selaku dosen ketua penguji yang telah memberikan saran dan arahan yang membangun dalam penulisan skripsi.
7. Angga Dwi Mulyanto, M.Si. selaku dosen anggota penguji 1 yang telah memberikan saran dan arahan yang membangun dalam penulisan skripsi.
8. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
9. Kedua orang tua tercinta, Ayah dan Ibu serta seluruh keluarga yang selalu memberikan doa, semangat, dan motivasi yang tiada henti kepada penulis.
10. Mahasiswa Program Studi Matematika angkatan 19, khususnya rekan-rekan KKM DR ~ Vacation Desa Tlogosari yang telah memberikan bantuan, semangat, dan motivasi kepada penulis.

11. Semua pihak yang memberikan semangat serta motivasi kepada penulis untuk segera menyelesaikan skripsi.

Semoga Allah Swt. memberikan balasan atas segala kebaikan yang telah diberikan kepada penulis. Penulis menyadari masih terdapat kekurangan dalam penyusunan laporan penelitian ini, Oleh karena itu, penulis mengharapkan saran dan kritik yang bersifat membangun dari semua pihak. Penulis berharap, semoga laporan penelitian ini dapat bermanfaat bagi pembaca serta dapat menambah ilmu pengetahuan.

Malang, 22 Juni 2023

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGAJUAN.....</b>	<b>ii</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>iii</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>iv</b>
<b>PERNYATAAN KEASLIAN TULISAN.....</b>	<b>v</b>
<b>MOTO.....</b>	<b>vi</b>
<b>PERSEMBAHAN .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xiv</b>
<b>ABSTRAK .....</b>	<b>xv</b>
<b>ABSTRACT .....</b>	<b>xvi</b>
<b>مستخلص البحث.....</b>	<b>xvii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	5
1.3 Tujuan Penelitian.....	5
1.4 Batasan Masalah .....	6
1.5 Manfaat Penelitian.....	6
<b>BAB II KAJIAN TEORI.....</b>	<b>7</b>
2.1 Fenomena Ketidakseimbangan ( <i>Imbalanced</i> ) .....	7
2.2 Variabel <i>Dummy</i> .....	8
2.3 <i>Logistic Regression</i> .....	9
2.3.1 <i>Logistic Regression</i> CERP .....	12
2.3.2 <i>Logistic Regression Ensemble</i> (LORENS).....	14
2.3.3 Evaluasi Performa Model.....	18
2.4 Kematian Bayi.....	19
2.5 Kematian Bayi dalam Kajian Islam.....	21
2.6 Kajian Kematian Bayi dengan Teori Pendukung.....	23
<b>BAB III METODE PENELITIAN.....</b>	<b>26</b>
3.1 Pendekatan Penelitian.....	26
3.2 Jenis dan Sumber Data .....	26
3.3 Variabel Penelitian .....	26
3.4 Tahapan Penelitian .....	28
3.5 Diagram Alir.....	29
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>30</b>
4.1 Statistik Deskriptif.....	30
4.1.1 Statistik Deskriptif Variabel Respon .....	30
4.1.2 Statistik Deskriptif Variabel Prediktor .....	32
4.2 Pembentukan Data <i>Dummy</i> .....	43
4.3 Analisis LORENS .....	44
4.3.1 Pembagian data <i>Training</i> dan <i>Testing</i> .....	44
4.3.2 Partisi dan <i>Ensemble</i> .....	45

4.3.3	Model LR pada Data <i>Training</i> .....	48
4.3.4	Nilai Prediksi .....	53
4.3.5	<i>Threshold</i> Optimal.....	54
4.3.6	Kelas Prediksi Klasifikasi.....	54
4.3.7	Menghitung Ketepatan Hasil Klasifikasi.....	55
4.4	Evaluasi Performa Model .....	57
4.5	Kajian Keislaman dengan Hasil Penelitian .....	60
<b>BAB V PENUTUP.....</b>		<b>63</b>
5.1	Kesimpulan.....	63
5.2	Saran .....	64
<b>DAFTAR PUSTAKA.....</b>		<b>65</b>
<b>LAMPIRAN.....</b>		<b>67</b>
<b>RIWAYAT HIDUP .....</b>		<b>78</b>

## DAFTAR TABEL

Tabel 2.1	Tabulasi Silang Klasifikasi <i>Actual</i> dan Klasifikasi Prediksi.....	16
Tabel 3.1	Variabel Angka Kematian Bayi .....	27
Tabel 4.1	<i>Imbalance Data</i> dan <i>Balance Data</i> pada Sensus Ibu Saat Melahirkan .....	33
Tabel 4.2	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Urutan Kelahiran.....	41
Tabel 4.3	Variabel Prediktor Hasil <i>Dummy</i> .....	44
Tabel 4.4	Lanjutan Variabel Prediktor Hasil <i>Dummy</i> .....	44
Tabel 4.5	Kombinasi Data <i>Training</i> dan Data <i>Testing</i> .....	45
Tabel 4.6	Kombinasi Partisi dan <i>Ensemble</i> .....	46
Tabel 4.7	Pengelompokkan Variabel Berdasarkan Partisi dan <i>Ensemble</i> ..	47
Tabel 4.8	Koefisien Model LR pada Setiap Partisi.....	48
Tabel 4.9	Koefisien Model LR pada Setiap Variabel .....	49
Tabel 4.10	Nilai Prediksi Setiap <i>Ensemble</i> .....	53
Tabel 4.11	Nilai Rata-Rata Probabilitas dan Keputusan Klasifikasi .....	55
Tabel 4.12	Hasil Tabulasi Silang Klasifikasi .....	56
Tabel 4.13	Partisi Optimal <i>Cross Validation</i> .....	58
Tabel 4.14	Hasil Threshold Optimal <i>Cross Validation</i> .....	58
Tabel 4.15	Tabulasi Silang Hasil Klasifikasi <i>Cross Validation</i> .....	59

## DAFTAR GAMBAR

Gambar 2.1	Bagan Konsep LR CERP.....	13
Gambar 2.2	Bagan Konsep LORENS .....	18
Gambar 3.1	Diagram Alir Penelitian.....	29
Gambar 4.1	<i>Imbalance Data</i> pada Sensus Kematian Bayi .....	31
Gambar 4.2	<i>Balance Data</i> pada Sensus Kematian Bayi .....	32
Gambar 4.3	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Berat Badan Bayi Lahir.....	33
Gambar 4.4	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Pendidikan Ibu.....	34
Gambar 4.5	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Pekerjaan Ibu .....	35
Gambar 4.6	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Interval Kelahiran.....	36
Gambar 4.7	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Frekuensi ANC.....	37
Gambar 4.8	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Tempat Persalinan .....	38
Gambar 4.9	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Jenis Kelamin Bayi.....	39
Gambar 4.10	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Pemberian ASI.....	40
Gambar 4.11	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Jenis Kelahiran .....	41
Gambar 4.12	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Tempat Tinggal .....	42
Gambar 4.13	<i>Imbalance Data</i> (i) dan <i>Balance Data</i> (ii) pada Sensus Indeks Kekayaan Keluarga .....	43

## DAFTAR LAMPIRAN

Lampiran 1. Data Angka Kematian Bayi 2017 .....	67
Lampiran 2. Data Hasil <i>Dummy</i> .....	68
Lampiran 3. Lanjutan Hasil Data <i>Dummy</i> .....	69
Lampiran 4. <i>Script</i> LORENS .....	70

## ABSTRAK

Munawar, Muhamad Sabit. 2023. **Implementasi *Logistic Regression Ensemble (LORENS)* dalam Pengklasifikasian Angka Kematian Bayi di Indonesia**. Skripsi, Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Ria Dhea Layla N.K., M.Si. (II) Mohammad Nafie Jauhari, M. Si.

**Kata Kunci:** AKB, *cross validation*, *ensemble*, LORENS, partisi, *threshold*

Angka Kematian Bayi (AKB) merupakan banyaknya kasus kematian bayi yang umurnya kurang dari satu tahun per 1000 kelahiran. AKB menjadi salah satu indikator terpenting dalam penentuan tingkat permasalahan kesehatan masyarakat. Banyaknya kasus AKB dipengaruhi beberapa faktor seperti pelayanan kesehatan, tingkat pendidikan orang tua, umur ibu saat melahirkan, dan masih banyak lagi. Tujuan dari penelitian ini adalah untuk memperoleh model LR dan tingkat ketepatan model pada pengklasifikasian AKB menggunakan *Logistic Regression Ensemble (LORENS)*. LORENS merupakan metode klasifikasi menggunakan teknik *ensemble* (penggabungan) yang dikembangkan berdasarkan metode *Logistic Regression*. Keunggulan LORENS yaitu bebas dari asumsi dimensi data dan penentuan kelas klasifikasinya menggunakan ambang (*threshold*) probabilitas yang optimal. Metode yang digunakan untuk mengevaluasi kebaikan model LORENS ini adalah *cross validation*. Akurasi terbaik terbentuk berdasarkan perbandingan data *training* dan data *testing* adalah 90%:10% dengan dua partisi dan 20 *ensemble*. Hasil klasifikasi AKB menggunakan LORENS membentuk 40 model LR dengan variabel pemberian ASI sebagai variabel paling berpengaruh terhadap kematian bayi. Hasil yang ditunjukkan pada perhitungan LORENS yaitu *sensitivity* adalah 76,45%, *specificity* adalah 84,08%, dan *accuracy* adalah 81,47%. Artinya klasifikasi AKB menggunakan metode LORENS menunjukkan hasil klasifikasi yang baik. Setelah dilakukan evaluasi terhadap metode LORENS menggunakan *cross validation* menunjukkan nilai *sensitivity* adalah 73,92%, *specificity* adalah 82,60%, dan *accuracy* adalah 79,64%.

## ***ABSTRACT***

Munawar, Muhamad Sabit. 2023. **Implementation of Logistic Regression Ensemble (LORENS) for classifying Infant Mortality Rates in Indonesia.** Undergraduate Thesis, Department of Mathematics, Faculty of Science and Thecnology. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Ria Dea Layla N.K., M.Si. (II) Mohammad Nafie Jauhari, M. Si.

**Keyword :** IMR, cross validation, ensemble, LORENS, partition, threshold

Infant Mortality Rate (IMR) is the number of cases of infant mortality less than one year old divided by 1000 births. IMR is one of the most important indicators in determining the level of public health problems. The number of IMR cases is influenced by several factors such as healthcare services, parental education level, mother's age at childbirth, and many more. The aim of this research is to obtain an LR model and the accuracy rate of the model in classifying IMR using Logistic Regression Ensemble (LORENS). LORENS is a classification method using ensemble techniques developed based on the Logistic Regression method. The advantage of LORENS is freedom from data dimension assumptions and the determination of classification classes using an optimal probability threshold. The method used to evaluate the goodness of the LORENS model is cross validation. The best accuracy is achieved based on a training data to testing data ratio of 90% : 10% with two partitions and 20 ensembles. The classification results of IMR using LORENS form 40 LR models, with breastfeeding variable being the most influential variable on infant mortality. The results shown in the LORENS calculation are sensitivity is 76,45%, the specificity is 84,08%, and the accuracy is 81,47%. This means that the classification of IMR using the LORENS method indicates good classification results. After evaluating the LORENS method using cross validation, the results show that the sensitivity is 73.92%, the specificity is 82.60%, and the accuracy is 79.64%.

## مستخلص البحث

منور, محمد ثابت. ٢٠٢٣. تنفيذ مجموعة الانحدار اللوجستي (*LORENS*) لتصنيف معدلات وفيات الرضع في إندونيسيا. البحث الجامعي، قسم الرياضيات ، كلية العلوم والتكنولوجيا ، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرفة: (١) ريا ديا ليلي نور كرسما, الماجستير (٢) محمد نافع جوهرى , الماجستير

الكلمات المفتاحية: *AKB*، التحقق المتقاطع ، المجموعة ، *LORENS*، التقسيم ، العتبة

معدل وفيات الرضع (*AKB*) هو عدد حالات وفيات الرضع أقل من سنة واحدة مقسوما على ١٠٠٠ ولادة. *AKB* هو واحد من أهم المؤشرات في تحديد مستوى مشاكل الصحة العامة. يتأثر عدد حالات *AKB* بعدة عوامل مثل خدمات الرعاية الصحية ومستوى تعليم الوالدين وعمر الأم عند الولادة وغيرها الكثير. الهدف من هذا البحث هو الحصول على نموذج الانحدار اللوجستي ومعدل دقة النموذج في تصنيف *AKB* باستخدام مجموعة الانحدار اللوجستي (*LORENS*). *LORENS* هو طريقة تصنيف باستخدام تقنيات المجموعة التي تم تطويرها بناء على طريقة الانحدار اللوجستي. ميزة *LORENS* هي التحرر من افتراضات أبعاد البيانات وتحديد فئات التصنيف باستخدام عتبة الاحتمال الأمثل. الطريقة المستخدمة لتقييم جودة نموذج *LORENS* هي التحقق المتبادل. يتم تحقيق أفضل دقة بناء على بيانات التدريب إلى نسبة بيانات الاختبار ٩٠٪: ١٠٪ مع قسمين و ٢٠ مجموعة. وتشكل نتائج تصنيف معدل وفيات الرضع باستخدام لورينز ٤٠ نموذجا من نماذج التحوف اللوجستي، مع كون متغير الرضاعة الطبيعية هو المتغير الأكثر تأثيرا على وفيات الرضع. النتائج الموضحة في حساب *LORENS* هي الحساسية ٧٦,٤٥٪، والنوعية ٨٤,٠٨٪، والدقة ٨١,٤٧٪. هذا يعني أن تصنيف *AKB* باستخدام طريقة *LORENS* يشير إلى نتائج تصنيف جيدة. بعد تقييم طريقة *LORENS* باستخدام التحقق المتقاطع ، أظهرت النتائج أن الحساسية هي ٧٣,٩٢٪، والنوعية ٨٢,٦٠٪، والدقة ٧٩,٦٤٪.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Tujuan program-program pemerintah hakekatnya adalah membangun kesejahteraan bagi rakyatnya. Kesejahteraan di suatu wilayah dapat dicapai dengan diadakannya pembangunan. Perbaikan dalam sektor masyarakat seperti sistem pendidikan, ekonomi, teknologi, infrastruktur, hingga kesehatan merupakan upaya-upaya dalam melakukan pembangunan. Salah satu komponen terpenting dalam penentuan tingkat kesejahteraan masyarakat adalah tingkat kesehatannya. Kesehatan masyarakat adalah tolak ukur dalam sebuah pembangunan, karena semakin baik tingkat kesehatan masyarakat, maka tingkat produktivitas pasti akan meningkat pula.

Mati dalam Kamus Besar Bahasa Indonesia memiliki arti sudah hilang nyawanya atau tidak hidup lagi. Kematian adalah berpisahny roh seseorang dari jasadnya yang menyebabkan jasad sudah tidak dapat berfungsi kembali. Kematian pasti terjadi kepada semua makhluk yang bernyawa. Hal tersebut tercantum dalam surah Al-Ankabut ayat 57 yang artinya sebagai berikut:

*“Tiap-tiap yang berjiwa akan merasakan mati. Kemudian hanyalah kepada Kami kamu dikembalikan”* (Kementerian Agama RI, 2019)

Menurut Tafsir Tahlili dalam Kementerian Agama RI (2019) ayat ini menerangkan bahwa tiap-tiap manusia pasti akan mati dan setelah mati, ia akan kembali kepada pemiliknya, yaitu Tuhan semesta alam. Sejak manusia dibangkitkan kembali di akhirat, sejak itu ia akan mengalami kehidupan yang sebenarnya dan selamanya. Bentuk kehidupan yang sebenarnya itu ditentukan

oleh sikap dan tindak-tanduk seseorang selama hidup di dunia. Jika ia seorang mukmin, maka akan memperoleh kebahagiaan yang abadi, sedangkan jika ia kafir, akan mengalami azab yang pedih di neraka. Ayat ini menjelaskan bahwa semua yang berjiwa akan merasakan mati, termasuk juga angka kematian bayi di dalamnya. Hal ini menunjukkan bahwa maut tidak memandang umur. Kapanpun dan di mana pun maut bisa datang kapan saja.

Menurut Alvaro dkk. (2019) dalam kajian Rencana Kerja Pemerintah (RKP) Anggaran Pendaatan dan Belanja Negara (APBN) DPR RI angka kematian bayi (AKB) adalah banyaknya kasus kematian bayi yang umurnya kurang dari satu tahun per 1000 kelahiran. AKB menjadi salah satu indikator terpenting dalam penentuan tingkat permasalahan kesehatan masyarakat. Tingginya tingkat AKB di suatu wilayah menggambarkan kurangnya pelayanan kesehatan di wilayah tersebut. Besarnya AKB dipengaruhi banyak faktor antara lain pelayanan kesehatan, tingkat pendidikan orang tua, umur ibu saat melahirkan, pekerjaan ibu, dan masih banyak lagi. Pada tahun 2017 nilai AKB yaitu 24, artinya dari 1000 kelahiran bayi hidup di Indonesia 24 diantaranya mengalami kematian. Kasus AKB tersebut jauh lebih rendah dari pada tahun 2012 yang sebanyak 32 bayi per 1000 kelahiran hidup. Namun angka tersebut belum memenuhi target *Millennium Development Goals* (MDGs) yaitu sebesar 23 pada 2015, dan target *Sustainable Development Goals* (SDGs) yaitu sebesar 12 pada 2030.

Klasifikasi merupakan metode multivariat yang memiliki hubungan dengan partisi sampel *training* dan alokasi pengamatan baru kedalam kelas atau kategori tertentu. Tujuannya adalah untuk memperoleh fungsi diskriminan optimal yang dapat memisahkan pengamatan yang berasal dari kelas yang berbeda, atau

memperoleh aturan yang bisa digunakan dalam penentuan kategori dari setiap pengamatan baru. Metode yang bisa digunakan dalam pengklasifikasian adalah *Logistic Regression* (LR), Analisis Diskriminan, *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), dan lain-lain. Di antara beberapa metode pengklasifikasian tersebut, LR adalah metode yang paling populer digunakan karena model dapat direpresentasikan dengan jelas dan ringkas (Widhianingsih dkk., 2020). Menurut Hosmer dkk. (2013) LR merupakan metode pendeskripsian hubungan antara variabel peubah respon (*dependent*) yang memiliki kategori sebanyak dua atau lebih dengan satu atau lebih variabel prediktor (*independent*) berskala kategori atau interval.

*Logistic Regression Ensemble* (LORENS) merupakan metode klasifikasi menggunakan teknik *ensemble* (penggabungan) yang dikembangkan oleh Lim pada 2010 (Ratnawati & Sunendiari, 2021). Metode klasifikasi ini merupakan pengembangan dari statistika dasar LR. Metode LORENS mengembangkan algoritma *Classifications by Ensembles from Random Partition* (CERP) dengan membagi menjadi beberapa subruang berdasarkan proses partisi dari varian-varianannya. Kemudian gabungkan kembali dalam model-model dari LR pada masing masing partisi kedalam satu fungsi (Asfihani, 2015).

Keunggulan LORENS apabila dibandingkan dengan LR yaitu bebas dari asumsi dimensi data, karena prediktor dalam LORENS dipartisi secara acak. Selain itu, LORENS menggunakan ambang (*threshold*) probabilitas pada masing-masing kelas yang optimal. Pada data dengan dimensi tinggi yang menggunakan LR akan menghasilkan permasalahan dikarenakan menggunakan ambang

(*threshold*) probabilitas 0,5. Hal tersebut menjadi masalah karena dirasa kurang adil jika probabilitas masing-masing kelas dinyatakan 0,5.

Penelitian sebelumnya dilakukan oleh T. Dwi Ary Widhianingsih dkk., yang berjudul *Logistic Regression Ensemble (LORENS) Applied to Drug Discovery*. Penelitiannya dilakukan untuk meningkatkan kinerja klasifikasi LR menggunakan LORENS. Subjek penelitian yang dilakukan terkait penemuan obat yang memiliki karakteristik data yang berdimensi tinggi. Model terbaik yang diperoleh dalam penelitian ini adalah menggunakan 25 partisi. Hasil eksperimen menunjukkan kinerja yang baik dengan akurasi sebesar 69,41% dan *Area Under Curve* (AUC) sebesar 0,7306 (Widhianingsih dkk., 2020). Penelitian lain dilakukan oleh Jainap Niken Melasasi pada 2015 yang berjudul *Klasifikasi Enzim pada Database DUD-E dengan Metode Logistic Regression Ensemble (LORENS)*. Penelitian ini dilakukan klasifikasi pada enzim database DUD-E yang meliputi enzim *aofh*, *cah2*, dan *hs90a*. Penelitian ini membagi data *training* 90% dan *testing* 10% dan diperoleh ketepatan klasifikasi sebesar 88,95 % untuk enzim *aofb* yang membagi data menjadi sembilan partisi optimal, 92,1% untuk enzim *cah2* yang membagi data menjadi empat partisi optimal, dan 100% untuk enzim *hs90a* yang membagi data menjadi lima partisi optimal (Melasasi, 2015). Penelitian lain dilakukan oleh Samira Ratnawati pada 2020 yang berjudul *penggunaan metode logistic regression ensemble (LORENS) pada klasifikasi leukimia akut*. Penelitian ini dilakukan untuk memperoleh hasil klasifikasi data kasus leukimia akut menggunakan LORENS. Data leukimia akut diolah dengan partisi optimal yaitu lima, *Threshold* optimal yaitu sebesar 0,4236, dan 15 *ensemble*. Hasil dari

penelitian ini adalah *sensitivity* sebesar 50%, *specificity* sebesar 83,33%, dan *accuracy* sebesar 78,57% (Ratnawati, 2020).

Berdasarkan uraian latar belakang yang ada, maka dilakukan penelitian terkait klasifikasi terhadap angka kematian bayi (AKB). Penentuan variabel diperoleh dari faktor yang mempengaruhi AKB hasil sensus kesehatan Badan Pusat Statistik (BPS). Penelitian ini nantinya akan dilakukan untuk memperoleh klasifikasi terbaik pada data AKB menggunakan metode LORENS.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang di atas diperoleh rumusan masalah sebagai berikut:

1. Bagaimana model yang terbentuk dari klasifikasi angka kematian bayi menggunakan LORENS?
2. Bagaimana hasil ketepatan klasifikasi angka kematian bayi menggunakan LORENS?

## **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah di atas diperoleh tujuan penelitian sebagai berikut:

1. Mengetahui model yang terbentuk dari klasifikasi angka kematian bayi menggunakan LORENS.
2. Memperoleh hasil ketepatan klasifikasi angka kematian bayi menggunakan LORENS.

#### 1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah

1. Data merupakan angka kematian bayi (AKB) di Indonesia pada tahun 2017.
2. Pengukuran tingkat *accuracy*, *sensitivity*, dan *specificity* menggunakan tabel tabulasi silang antara kelas *actual* dan kelas prediksi.
3. Evaluasi Performa Model yang digunakan adalah menggunakan *Cross Validation*

#### 1.5 Manfaat Penelitian

Manfaat yang diharapkan penelitian ini adalah

1. Bagi Pembaca  
Memberikan informasi terkait performansi metode *Logistic Regression Ensemble* (LORENS) sebagai metode pengklasifikasian.
2. Bagi Peneliti Selanjutnya  
Memberi tambahan referensi dan bahan bacaan kepada penelitian selanjutnya yang akan mengkaji terkait pengklasifikasian menggunakan *Logistic Regression Ensemble* (LORENS).
3. Bagi Instansi  
Model klasifikasi dapat digunakan sebagai gambaran dan bahan evaluasi angka kematian bayi (AKB) di Indonesia, sehingga bisa dijadikan bahan edukasi untuk mengurangi angka kematian bayi di masa depan.

## BAB II

### KAJIAN TEORI

#### 2.1 Fenomena Ketidakseimbangan (*Imbalanced*)

Sebuah kumpulan data dikatakan tidak seimbang atau *imbalanced* jika jumlah pada suatu kelas tertentu (kelas mayor) jauh lebih banyak dibandingkan dengan jumlah kelas yang lain (kelas minor). Keadaan ini menjadi salah satu keadaan yang krusial untuk ditangani dalam kasus-kasus data mining, karena dapat mempengaruhi nilai akurasi saat proses klasifikasi data (Choirunnisa, 2019).

Pendekatan yang sering dilakukan untuk mengatasi keadaan *imbalanced* ini adalah pendekatan pada proses *sampling*. Pendekatan ini dilakukan dengan melakukan *sampling* pada data mayoritas maupun data minoritas sehingga jumlah data lebih seimbang. Umumnya pada teknik *sampling* yang dilakukan terdapat dua jenis yaitu *undersampling*, dan *oversampling*. *Under sampling* adalah metode *sampling* yang dilakukan dengan mengurangi atau mengeliminasi Sebagian data pada kelas mayoritas pada data. Sedangkan *over sampling* adalah metode *sampling* yang dilakukan dengan menambahkan jumlah data pada kelas minoritas sehingga diharapkan rasio antarkelas minoritas dan kelas mayoritas lebih seimbang.

Penelitian ini menggunakan *over sampling* untuk mengatasi keadaan *imbalanced* pada data. Secara umum permasalahan pada *over sampling* adalah terjadinya *over fitting* dikarenakan penambahan data yang berulang menyebabkan *decision boundary* menjadi lebih ketat. Oleh karenanya, pada perkembangannya metode *over sampling* bukan lagi dilakukan dengan menyalin data yang sama,

namun membuat data baru yang mirip. Hal ini bertujuan untuk memperhalus *decision boundary* (Choirunnisa, 2019).

## 2.2 Variabel *Dummy*

Analisis regresi seringkali terdiri dari beberapa jenis data, antara lain rasio berupa pendapatan, harga, biaya, tinggi, dan lainnya. Selain itu juga jenis data nominal seperti jenis kelamin, ras, warna kulit, agama, dan lainnya. Variabel nominal umumnya merupakan variabel kualitatif, sehingga perlu diperjelas. Salah satu cara untuk menjelaskan terhadap atribut tersebut adalah dengan membangun bantuan yang bernilai 1 dan 0 yang merupakan tanda kepemilikan dari suatu atribut. Nilai 1 untuk menandakan adanya atribut dan 0 menandakan tidak adanya atribut. Hubungan variabel kualitatif dapat dilihat dengan variabel bebas, maka metode yang digunakan adalah metode variabel *dummy*. Variabel *dummy* merupakan variabel yang sudah dikodekan. Jika data kualitatif memiliki  $m$  kategori, maka jumlah *dummy* variabel yang dicantumkan dalam model adalah  $(m - 1)$  dikarenakan agar tidak terjadinya jebakan *dummy* yaitu di mana situasi kolinearitas sempurna jika terdapat lebih dari satu hubungan (Baudi, 2019).

Misal data merupakan data kategorik dengan 3 kategori, yaitu kategori a, kategori b, dan kategori c. Jika kategori a sebagai kategori dasar, maka akan diperoleh 2 variabel *dummy*, yaitu *dummy* 1 ( $D_1$ ) yaitu bernilai 1 untuk kategori b dan bernilai 0 untuk yang lainnya. Serta *dummy* 2 ( $D_2$ ) yaitu bernilai 1 untuk kategori c dan bernilai 0 untuk yang lainnya.

### 2.3 *Logistic Regression*

*Logistic regression* (LR) merupakan suatu metode analisis statistika untuk menganalisis model terbaik berdasarkan hasil hubungan yang terjadi antara variabel prediktor (*independent*) dan variabel peubah respon (*dependent*). Variabel peubah respon memiliki dua kategori atau lebih, sedangkan variabel prediktor berskala kategori maupun interval yang memiliki satu atau lebih variabel (Hosmer & Lemeshow, 2000). LR ialah regresi *non-linear* yang digunakan untuk mengetahui hubungan antara peubah prediktor dan peubah respon yang memiliki sifat tidak linier. Ketidaknormalan sebaran  $y$  dan keragaman respon tidak konstan yang tidak bisa dijelaskan dengan model regresi linier biasa (Agresti, 2013).

Variabel respon ( $y$ ) dalam *binary logistic regression* memiliki dua kategori yaitu “gagal” dan “sukses”. Variabel respon kategori “gagal” dinotasikan dengan  $y = 0$  dan kategori “sukses” dinotasikan dengan  $y = 1$ . Artinya variabel respon  $y$  dapat diasumsikan mengikuti distribusi Bernoulli dalam setiap observasi tunggalnya. Fungsi probabilitas dalam setiap observasi adalah

$$f(y_i) = (\pi_j(x_i))^{y_i} (1 - \pi_j(x_i))^{1-y_i}; y_i = 0,1 \quad (2.1)$$

dengan  $\pi_j(x_i)$  adalah probabilitas dari kejadian ke  $-j$  pada data ke- $i$ . Apabila  $y_i = 1$ , maka  $f(y_i) = \pi_j(x_i)$  dan apabila  $y_i = 0$ , maka  $f(y_i) = 1 - \pi_j(x_i)$ .

Berdasarkan fungsi probabilitas pada (2.1) diperoleh fungsi *Logistic Regression* (LR) sebagai berikut:

$$f(z) = \frac{e^z}{1 + e^z} \quad (2.2)$$

dengan  $z$  adalah  $\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$  dan  $p$  adalah banyak variabel prediktor.

Setiap nilai  $z$  yang diberikan, maka nilai  $f(z)$  terletak diantara 0 dan 1. Karena  $z$  adalah nilai yang terletak di antara  $-\infty$  dan  $\infty$ . Hal tersebut menunjukkan bahwa model LR sebenarnya mengilustrasikan resiko atau probabilitas dari suatu objek. Sehingga Persamaan 2.2 dapat dituliskan sebagai berikut.

$$\pi_j(x_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}} \quad (2.3)$$

Perhatikan bahwa hubungan antara variabel prediktor dan variabel respon dalam LR merupakan suatu fungsi non-linier. Sehingga bentuk Persamaan 2.3 dapat ditransformasi kedalam bentuk linier atau dalam regresi logistik disebut transformasi logit dari  $\pi_j(x_i)$  sebagai berikut.

$$\text{Logit}[\pi_j(x_i)] = \ln\left(\frac{\pi_j(x_i)}{1 - \pi_j(x_i)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (2.4)$$

Sedangkan estimasi parameter dalam LR dapat menggunakan Metode *Maximum Likelihood* (MLE). Metode ini memberikan nilai estimasi  $\beta$  dengan memaksimalkan fungsi *likelihood* (Hosmer & Lemeshow, 2000). Syarat Metode MLE adalah data harus didasarkan pada suatu distribusi tertentu. Pada LR setiap pengamatan didasarkan pada distribusi Bernoulli, sehingga fungsi *likelihood* dapat ditentukan dengan persamaan berikut.

$$\begin{aligned} L(\boldsymbol{\beta}, \mathbf{X}) &= \prod_{j,i=1}^n [\pi_j(x_i)]^{y_i} [1 - \pi_j(x_i)]^{1-y_i} \quad (2.5) \\ &= \left\{ \prod_{j,i=1}^n \pi_j(x_i)^{y_i} \right\} \left\{ \prod_{j,i=1}^n (1 - \pi_j(x_i)) (1 - \pi_j(x_i))^{-y_i} \right\} \\ &= \left\{ \prod_{j,i=1}^n (1 - \pi_j(x_i)) \right\} \left\{ \prod_{j,i=1}^n \pi_j(x_i)^{y_i} \right\} \left\{ \prod_{j,i=1}^n (1 - \pi_j(x_i))^{-y_i} \right\} \end{aligned}$$

$$\begin{aligned}
&= \left\{ \prod_{j,i=1}^n (1 - \pi_j(x_i)) \right\} \left\{ \prod_{j,i=1}^n e^{\left( \ln \left( \frac{\pi_j(x_i)}{1 - \pi_j(x_i)} \right)^{y_i} \right)} \right\} \\
&= \left\{ \prod_{j,i=1}^n (1 - \pi_j(x_i)) \right\} \left\{ e^{\left( \sum_{i=1}^n y_i \ln \left( \frac{\pi_j(x_i)}{1 - \pi_j(x_i)} \right) \right)} \right\} \\
&= \left\{ \prod_{j,i=1}^n \left( \frac{1}{1 + e^{\sum_{p=0}^m \beta_p x_{ip}}} \right) \right\} \left\{ e^{\left( \sum_{p=0}^m \beta_p \left( \sum_{i=1}^n y_i x_{ip} \right) \right)} \right\}
\end{aligned}$$

MLE diperoleh dengan memaksimalkan logaritma fungsi *likelihood* pada Persamaan 2.5 sehingga diperoleh

$$\begin{aligned}
\ln L(\boldsymbol{\beta}, \mathbf{X}) &= \ln \left[ \left\{ \prod_{j,i=1}^n \left( \frac{1}{1 + e^{\sum_{p=0}^m \beta_p x_{ip}}} \right) \right\} \left\{ e^{\left( \sum_{p=0}^m \beta_p \left( \sum_{i=1}^n y_i x_{ip} \right) \right)} \right\} \right] \quad (2.6) \\
&= \left\{ \ln \left( e^{\left( \sum_{p=0}^m \beta_p \left( \sum_{i=1}^n y_i x_{ip} \right) \right)} \right) \right\} \left\{ \ln \left( \prod_{j,i=1}^n (1 + e^{\sum_{p=0}^m \beta_p x_{ip}})^{-1} \right) \right\} \\
&= \sum_{p=0}^m \beta_p \left( \sum_{j,i=1}^n y_i x_{ip} \right) - \sum_{j,i=1}^n \ln(1 + e^{\sum_{p=0}^m \beta_p x_{ip}})
\end{aligned}$$

nilai  $\boldsymbol{\beta}$  diperoleh dengan membuat turunan pertamanya

$$\begin{aligned}
\frac{\partial \ln L(\boldsymbol{\beta}, \mathbf{X})}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left[ \sum_{p=0}^m \beta_p \left( \sum_{j,i=1}^n y_i x_{ij} \right) - \sum_{j,i=1}^n \ln(1 + e^{\sum_{p=0}^m \beta_p x_{ip}}) \right] \quad (2.7) \\
&= \sum_{j,i=1}^n y_i x_{ij} - \sum_{j,i=1}^n x_{ij} \left( \frac{e^{\sum_{p=0}^m \beta_p x_{ip}}}{1 + e^{\sum_{p=0}^m \beta_p x_{ip}}} \right) \\
&= \sum_{j,i=1}^n y_i x_{ij} - \sum_{j,i=1}^n x_{ij} \pi_j(x_i) \\
&= \sum_{j,i=1}^n x_{ij} (y_i - \pi_j(x_i)) \\
&= \mathbf{X}^T (\mathbf{y} - \hat{\boldsymbol{\pi}})
\end{aligned}$$

dengan  $\mathbf{y}$  adalah vektor pengamatan pada variabel respon yang berukuran  $n \times 1$ , sedangkan  $\mathbf{X}$  adalah matriks variabel prediktor yang berukuran  $n \times (p + 1)$ .

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & a_{1p} \\ 1 & x_{21} & x_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.8)$$

Jika  $\frac{\partial \ln L(\boldsymbol{\beta}, \mathbf{X})}{\partial \boldsymbol{\beta}} = 0$  dan  $\hat{\mathbf{y}} = \hat{\boldsymbol{\pi}}$  maka didapatkan persamaan (2.8) sebagai berikut.

$$\mathbf{X}^T (\mathbf{y} - \hat{\mathbf{y}}) = 0 \quad (2.9)$$

Persamaan (2.9) dapat diselesaikan menggunakan metode newton-Raphson. sehingga diperoleh hasil estimasi parameter  $\hat{\boldsymbol{\beta}}$  sebagai berikut:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{z} \quad (2.10)$$

di mana  $\mathbf{z}$  adalah vektor berukuran  $n \times 1$  dan  $\mathbf{W}$  adalah vektor pembobot, dengan  $\mathbf{z}$  sebagai berikut:

$$\mathbf{z} = \text{logit}[\hat{\boldsymbol{\pi}}(\mathbf{x})] + \frac{\mathbf{y} - \hat{\boldsymbol{\pi}}(\mathbf{x})}{\hat{\boldsymbol{\pi}}(\mathbf{x})[1 - \hat{\boldsymbol{\pi}}(\mathbf{x})]} \quad (2.11)$$

serta matriks varian kovarian untuk  $\hat{\boldsymbol{\beta}}$  ditampilkan dengan persamaan berikut.

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \text{diag}[\hat{\boldsymbol{\pi}}(1 - \hat{\boldsymbol{\pi}})] \mathbf{X})^{-1} \quad (2.12)$$

### 2.3.1 Logistic Regression CERP

*Logistic Regression Classification by Ensembles from Random Partitions* (LR CERP) merupakan pengklasifikasian menggunakan *ensemble* di mana LR sebagai basis pengklasifikasian. Ruang variabel prediktor dalam algoritma LR CERP dipartisi secara acak menjadi beberapa subruang yang saling *mutually exclusive*. Misalnya dalam sebuah ruang prediktor  $\Theta$  dipartisi menjadi beberapa subruang  $(\theta_1, \theta_2, \dots, \theta_n)$  yang saling *mutually exclusive* dengan ukuran yang sama. Sehingga pada masing-masing subruang dapat diasumsikan tidak terdapat bias (Melasasi, 2015).

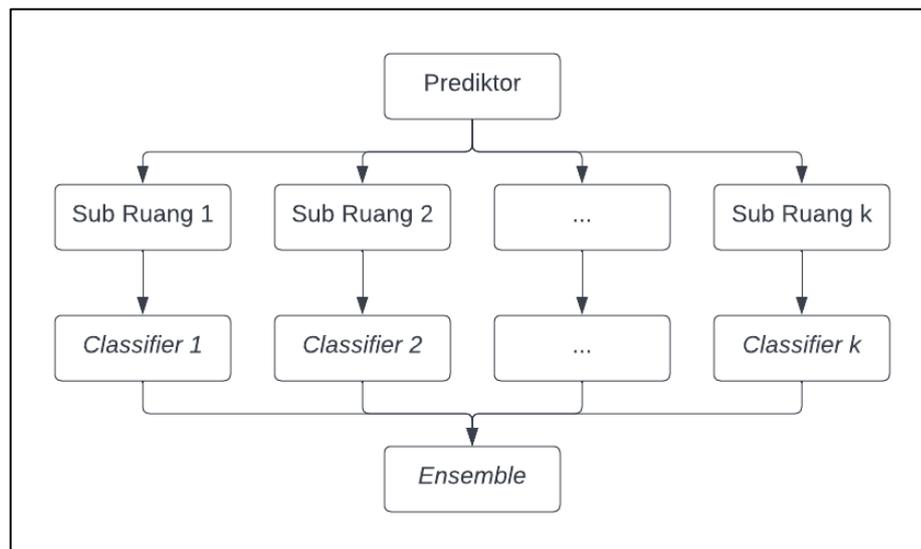
Performa LR CERP sesuai dengan banyaknya variabel prediktor di setiap partisi. Sebelum menentukan banyak variabel dalam partisi, perlu ditentukan jumlah partisi paling optimal. Penentuannya sendiri di dipengaruhi banyak data. Data yang memiliki jumlah data ( $n$ ) lebih besar dibanding banyak variabel prediktor ( $p$ ) dapat diperoleh partisi menggunakan persamaan berikut:

$$K = \frac{p}{q} \quad (2.13)$$

di mana  $q$  adalah bilangan integer yang kurang dari  $n$ . Sedangkan untuk data yang memiliki jumlah data ( $n$ ) lebih kecil dibanding banyak variabel prediktor ( $p$ ) dapat diperoleh partisi paling optimal menggunakan persamaan berikut.

$$K = \frac{6 \times p}{n} \quad (2.14)$$

Penjelasan LR CERP dapat digambarkan dalam bagan konsep sebagai berikut.



Gambar 2.1 Bagan Konsep LR CERP

Pembentukan model klasifikasi akan dibuat pada setiap subruang menggunakan LR, namun model LR lemah terhadap pemilihan variabel (Lim,

2007). LR CERP digunakan untuk peningkatan akurasi di mana hasil klasifikasi pada setiap subruang yang dibentuk akan dikombinasikan. Kombinasi beberapa model LR yang dilakukan pada LR CERP dengan mengambil rata-rata nilai prediksi yang dihasilkan setiap *ensemble* digunakan untuk meningkatkan akurasi. Nilai prediksi yang dihasilkan dari semua hasil klasifikasi dalam subruang dirata-rata dan diklasifikasikan sebagai 1 atau 0 berdasarkan *threshold* (Ahn dkk., 2007).

### 2.3.2 *Logistic Regression Ensemble (LORENS)*

*Logistic Regression Ensemble (LORENS)* merupakan metode klasifikasi yang dikembangkan oleh Lim dkk. pada tahun 2010. Dasar klasifikasi LORENS adalah menggunakan model LR yang dikembangkan berdasarkan algoritma LR CERP. Ketika hasil model LR dikombinasikan pada LORENS untuk mendapatkan satu klasifikasi yang kuat dibandingkan dengan metode agregasi kompleks lainnya. LORENS membentuk beberapa *ensemble* dengan mengulangi prosedur LR CERP beberapa kali (Widhianingsih, 2018).

Sama halnya dengan LR CERP, LORENS mempartisi data kedalam  $k$  subruang yang ditentukan dengan cara acak dan distribusi yang sama. Sehingga dapat diasumsikan bahwa pemilihan variabel tidak mengalami bias di setiap subruang. Pada setiap subruang menggunakan model LR tanpa pemilihan variabel. Kemudian dari partisi variabel yang dilakukan secara acak tersebut diharapkan akan diperoleh probabilitas *error* yang hampir sama dari  $k$  klasifikasi, sehingga dapat meningkatkan akurasi. LORENS dapat meningkatkan akurasi dalam setiap *ensemble* yang dihasilkan. Peningkatan

akurasi dalam satu *ensemble* diperoleh dari hasil kombinasi nilai prediksi masing-masing model LR pada setiap partisi. Pengulangan prosedur LR CERP pada LORENS menghasilkan kombinasi rata-rata atau nilai terbanyak dengan akurasi yang hampir sama. LORENS menggunakan hasil rata-rata dengan nilai sedikit yang lebih unggul dibandingkan dengan nilai terbanyak.

Beberapa *ensemble* yang dihasilkan LORENS memiliki partisi yang berbeda-beda sesuai nilai terbanyak dari setiap *ensemble*. Berdasarkan nilai tersebut diperoleh satu akurasi umum yang akurasinya telah ditingkatkan. Umumnya, peningkatan akurasi ini akan diperoleh jika jumlah *ensemble* yang dibangun lebih dari 10. Proses klasifikasi dalam LR umumnya berdasarkan ambang (*threshold*) probabilitas. *Threshold* yang umum digunakan dalam klasifikasi adalah 0.5, namun akurasi klasifikasi tidak akan baik apabila proporsi kelas 0 dan kelas 1 tidak seimbang. Sebagai penyeimbangan *sensitivity* dan *specificity*, LORENS menggunakan *threshold* yang optimal. *Threshold* tersebut diperoleh menggunakan rumus berikut (Lim dkk., 2010):

$$Threshold = \frac{\bar{y} + 0,5}{2} \quad (2.15)$$

di mana  $\bar{y}$  merupakan proporsi respon positif yang terdapat pada data.

Proses klasifikasi dalam metode LORENS, perlu dilakukan tahapan-tahapan berikut:

1. Membentuk model logit berdasarkan data *training*;
2. Memastikan data *testing* kedalam model logit, sehingga didapatkan nilai prediksi;
3. Klasifikasi pengamatan pada data *testing*, apabila nilai probabilitas yang diperoleh lebih kecil daripada nilai *threshold* maka pengamatan masuk

kedalam kelas negatif, sebaliknya jika probabilitas yang diperoleh lebih besar daripada nilai *threshold* maka pengamatan masuk kedalam kelas positif;

4. Bandingkan hasil pada kelas *actual* dengan kelas prediksi klasifikasi;
5. Kelompokkan hasil perbandingan kedalam kelompok *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

*True Positive* (TP) adalah total kematian bayi kelas positif yang terprediksi di kelas positif, *True Negative* (TN) adalah total kematian bayi kelas negatif yang terprediksi di kelas negatif, *False Positive* (FP) adalah total kematian bayi kelas negatif yang terprediksi menjadi positif, dan *False Negative* (FN) adalah total kematian bayi kelas positif yang terprediksi menjadi negatif.

Tabel 2.1 Tabulasi Silang Klasifikasi *Actual* dan Klasifikasi Prediksi

	<b>Kelas <i>actual</i></b>		
	$p(+)$	$n(-)$	
<b>Kelas Prediksi</b>	$p(+)$	TP	FP
	$n(-)$	FN	TN

Ketepatan prediksi dalam klasifikasi dapat diperoleh dengan membagi jumlah prediksi yang tepat dengan total jumlah prediksi. Berikut merupakan rumus yang digunakan untuk memperoleh ketepatan klasifikasi berupa *sensitivity*, *specificity*, dan *accuracy* (Catal, 2012).

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (2.16)$$

$$Specificity = \frac{TN}{(FP + TN)} \quad (2.17)$$

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (2.18)$$

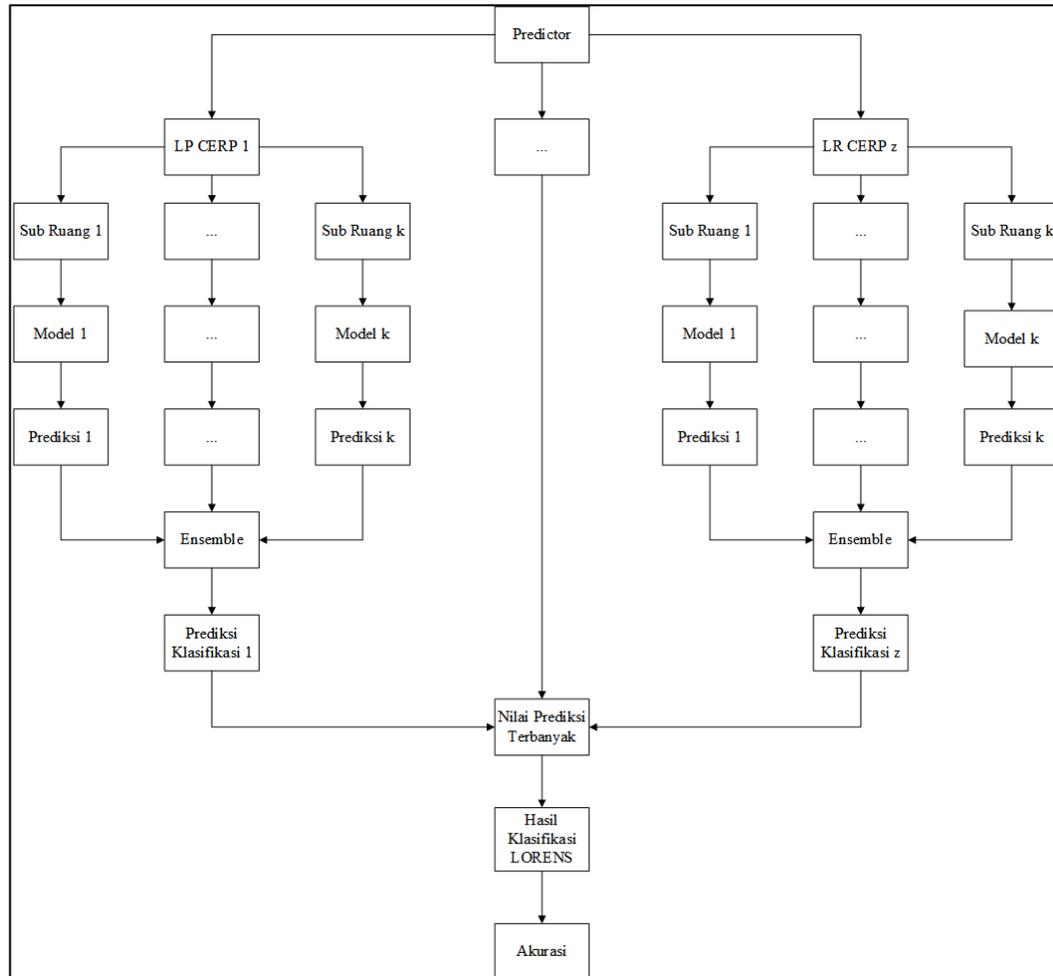
Sedangkan kesalahan dalam klasifikasi dapat diperoleh dengan rating kesalahan positif atau *False Positive Rate* (FPR), rating kesalahan negatif atau *False Negative Rate* (FNR), dan *Error*. Berikut merupakan rumus untuk menghitung kesalahan dalam klasifikasi.

$$FPR = \frac{FP}{(TP + FN)} \quad (2.19)$$

$$FNR = \frac{FN}{(TP + FN)} \quad (2.20)$$

$$Error = \frac{(FP + FN)}{(TP + FP + TN + FN)} \quad (2.21)$$

Beberapa *ensemble* dihasilkan metode LORENS dan memilih nilai prediksi terbanyak diantara *ensemble*. Berdasarkan nilai prediksi tersebut, diperoleh satu akurasi yang nilainya telah ditingkatkan berdasarkan beberapa *ensemble*. LORENS dalam model klasifikasi memiliki kelebihan yaitu bebas dari asumsi dimensi data, dikarenakan LORENS mempartisi acak variabel-variabel prediktornya. Dibandingkan dengan LR CERP, LORENS lebih unggul dalam basis klasifikasi menggunakan LR yang populer dan mudah dipahami (Lim dkk., 2010). Gambar 2.2 merupakan gambaran atau bagan konsep dari metode LORENS.



Gambar 2.2 Bagan Konsep LORENS

### 2.3.3 Evaluasi Performa Model

Metode yang digunakan untuk mengevaluasi performa model adalah *cross validation*. Metode *cross validation* adalah metode untuk mengevaluasi performa model dengan membagi data menjadi  $k$  *folds* atau partisi yang memiliki jumlah seimbang. Berdasarkan partisi tersebut, *cross validation* juga membagi data menjadi data *testing* dan data *training*. Semua partisi digunakan sebagai data *testing* sekaligus data *training*. Penggunaan data sebagai data *testing* atau data *training* dilakukan saat gilirannya masing-masing. Giliran masing-masing data dapat dilakukan karena prosedur berulang dari *cross*

*validation*. Misalkan dalam *cross validation* digunakan sebagai data *testing* sebanyak satu partisi,  $k - 1$  sisanya digunakan sebagai data *training*. Selanjutnya, dilakukan secara berulang hingga semua partisi telah dijadikan sebagai data *testing*. Prosedur ini sebut dengan *k fold cross validation* (Witten dkk., 2011).

Misalnya metode *cross validation* yang digunakan 10 partisi. Pertama bagi data secara acak menjadi 10 partisi yang seimbang. Lakukan metode *cross validation* dengan 1 partisi sebagai data *testing* dan 9 partisi sisanya sebagai data *training*. Lakukan sebanyak 10 kali dengan masing-masing partisi telah menjadi data *testing* dan data *training*. Beberapa penelitian menyebutkan bahwa 10 partisi dalam *cross validation* menghasilkan nilai prediksi terbaik. *Cross validation* dengan 10 partisi ini dapat disebut *10 folds cross validation*. Metode *10 folds cross validation* telah menjadi metode standar dalam *machine learning* dan data *mining* (Melasasi, 2015).

## 2.4 Kematian Bayi

Menurut Alvaro dkk. (2019) dalam kajian RKP APBN DPR RI Kematian bayi merupakan kematian yang terjadi pada bayi setelah bayi lahir hingga bayi berumur kurang dari satu tahun. Angka kematian bayi (AKB) merupakan banyaknya kasus kematian bayi yang terjadi per 1000 kelahiran. Penentuan tingkat permasalahan kesehatan masyarakat dapat dilihat dari tingkat AKB. AKB menjadi salah satu indikator yang dicanangkan sistem kesehatan nasional dalam mewujudkan keberhasilan pembangunan kesehatan, bahkan dipakai sebagai indikator sentral keberhasilan kesehatan di Indonesia.

Penyebab kematian bayi dibagi menjadi dua macam yaitu eksogen dan endogen. Kematian bayi eksogen adalah kematian pada bayi yang disebabkan oleh faktor-faktor yang berkaitan dengan lingkungan luar seperti tingkat pengetahuan seorang ibu atau bahkan kekurangan gizi karena faktor ekonomi. Sedangkan endogen adalah kematian pada bayi yang disebabkan oleh faktor-faktor bawaan bayi sejak lahir seperti bayi lahir prematur atau kelainan *kongenital* (Azizah, 2013).

Banyak faktor yang mempengaruhi kematian bayi seperti kurangnya kesadaran ibu atas kesehatan, ibu jarang memeriksa kandungannya ke bidan, hamil di usia muda, kurang asupan gizi bagi ibu dan bayinya, hamil di usia tua, makanan yang dikonsumsi kurang higienis, dan masih banyak lagi. Selain itu kondisi ibu saat hamil juga mempengaruhi kesehatan kandungannya, seperti faktor psikologis, fisik, sosial, dan budaya (Sulistyawati, 2009).

Masyarakat internasional telah berkomitmen untuk mempercepat pengentasan kemiskinan dan pembangunan manusia dengan adanya *Millenium Development Goals* (MDGs). Tujuan MDGs yaitu menurunkan angka kematian balita sebesar dua pertiga dari tahun 1990 sampai dengan 2015. AKB adalah indikator terpenting dalam angka kematian balita. Target MDGs terkait AKB adalah 23 kematian per 1000 kelahiran. Sedangkan AKB di Indonesia pada tahun 2017 yaitu sebanyak 24 kematian per 1000 kelahiran. Artinya Indonesia belum bisa mencapai target yang ditetapkan MDGs. Indonesia masih harus terus berbenah untuk menanggulangi permasalahan kesehatan untuk mencapai target MDGs (Alvaro dkk., 2019).

## 2.5 Kematian Bayi dalam Kajian Islam

Kematian dalam Islam adalah berpisahannya roh seseorang dari jasadnya yang menyebabkan jasad sudah tidak dapat berfungsi kembali. Kematian pasti terjadi kepada semua makhluk yang bernyawa. Kematian dapat datang di waktu yang tidak diketahui oleh manusia. Terdapat seseorang yang mati ketika ia berusia sudah dewasa hingga lansia. Tidak sedikit pula yang mati ketika berusia anak-anak, bayi, dan saat masih dalam kandungan. Lama hidup seseorang adalah semata-mata kehendak dan kuasa dari Allah SWT. Hal tersebut tercantum dalam surah Al-Hajj ayat 5 yang artinya sebagai berikut:

*“Hai manusia, jika kamu dalam keraguan tentang kebangkitan (dari kubur), maka (ketahuilah) sesungguhnya Kami telah menjadikan kamu dari tanah, kemudian dari setetes mani, kemudian dari segumpal darah, kemudian dari segumpal daging yang sempurna kejadiannya dan yang tidak sempurna, agar Kami jelaskan kepada kamu dan Kami tetapkan dalam rahim, apa yang Kami kehendaki sampai waktu yang sudah ditentukan, kemudian Kami keluarkan kamu sebagai bayi, kemudian (dengan berangsur-angsur) kamu sampailah kepada kedewasaan, dan di antara kamu ada yang diwafatkan dan (adapula) di antara kamu yang dipanjangkan umurnya sampai pikun, supaya dia tidak mengetahui lagi sesuatupun yang dahulunya telah diketahuinya. Dan kamu lihat bumi ini kering, kemudian apabila telah Kami turunkan air di atasnya, hiduplah bumi itu dan suburlah dan menumbuhkan berbagai macam tumbuh-tumbuhan yang indah”* (Kementerian Agama RI, 2019)

Menurut Tafsir Wajiz Kementerian Agama RI (2019), surah Al-Hajj ayat 5 menerangkan jawaban atas orang-orang yang meragukan hari kebangkitan dengan memperhatikan perkembangan hidup manusia. Allah telah menciptakan manusia dari tanah. Kemudian bercampurnya setetes mani dengan sel telur yang menjadi segumpal darah dan berkembang menjadi segumpal daging. Dari segumpal daging tersebut disempurnakan pertumbuhannya menjadi manusia tanpa cacat apa pun, dan ada pula yang tidak sempurna, karena cacat fisik maupun mental sejak dari kandungan. Supaya Allah jelaskan kepada manusia bahwa manusia berada dalam

kekuasaan-Nya. Allah tetapkan manusia sewaktu embrio dalam rahim seorang ibu menurut kehendak-Nya hingga tiap orang berbeda rentang waktu berada dalam kandungan ibunya sampai waktu yang sudah ditentukan, biasanya setelah 36 minggu. Kemudian Allah keluarkan manusia dari rahim ibunya sebagai bayi, kemudian dengan berangsur-angsur kamu tumbuh-kembang sampai kepada usia dewasa. Diantara bayi tersebut ada yang diwafatkan pada usia bayi, muda, bahkan ada pula yang diberi umur panjang, serta dikembalikan pada usia pikun karena sangat tua.

Berdasarkan surah Al-Hajj ayat 5 jelas bahwa umur atau usia seseorang untuk hidup di dunia adalah rahasia-Nya. Tidak seorangpun tau kapan ia akan mati. Terdapat orang yang hidup hingga umur tuanya, terdapat pula orang yang hidup hingga masa dewasanya saja. Namun tidak sedikit orang yang mati di masa bayinya.

Pada hakikatnya bayi yang dilahirkan ke dunia dalam keadaan fitrah Islam, orang sekitarnya lah yang membuat bayi tersebut menjadi beragama selain Islam. Hal tersebut tercantum dalam hadis Rasulullah yang dikutip dari hadis shahih Al-Bukhari No. 4402 yang artinya sebagai berikut.

*“Seorang bayi tidak dilahirkan (ke dunia ini) melainkan ia berada dalam kesucian (fitrah). Kemudian kedua orang tuanyalah yang akan membuatnya menjadi Yahudi, Nasrani, ataupun Majusi”* (HR. Bukhari)

Selain itu, Rasulullah juga pernah bersabda terkait tiga golongan orang yang di angkat pena atau catatan amal darinya dalam hadis yang dikutip dari hadis Sunan Ibnu Majah No. 2032 berikut yang artinya sebagai berikut.

*“Pena diangkat dari (tiga) golongan; anak kecil, orang gila dan orang tidur”* (HR. Ibnu Majah)

Kedua Hadis di atas menjelaskan bahwa bayi lahir dalam keadaan fitrah dan hingga sebelum bayi tersebut menjapai umur baligh, maka pena yang menulis segala amal di angkat darinya. Oleh karenanya, anak kecil khususnya bayi tetap dalam keadaan fitrah hingga ia mencapai baligh. Sejatinya manusia yang mati dalam keadaan fitrah sudah pasti masuk kedalam surga.

## **2.6 Kajian Kematian Bayi dengan Teori Pendukung**

Kematian bayi merupakan kematian yang terjadi pada bayi setelah bayi lahir hingga bayi berumur kurang dari satu tahun. Angka kematian bayi (AKB) merupakan banyaknya kasus kematian bayi yang terjadi per 1000 kelahiran. Penentuan tingkat permasalahan kesehatan masyarakat dapat dilihat dari tingkat AKB. AKB menjadi salah satu indikator yang dicanangkan sistem kesehatan nasional dalam mewujudkan keberhasilan pembangunan kesehatan, bahkan dipakai sebagai indikator sentral keberhasilan Kesehatan di Indonesia.

Masyarakat internasional telah berkomitmen untuk mempercepat pengentasan kemiskinan dan pembangunan manusia berupa *Millenium Development Goals* (MDGs). Salah satu tujuan MDGs ini adalah untuk menurunkan angka kematian balita. Angka kematian bayi menjadi indikator terpenting dalam menurunkan angka kematian balita. Target MDGs terkait AKB adalah 23 kematian per 1000 kelahiran. Sedangkan AKB di Indonesia pada tahun 2017 yaitu sebanyak 24 kematian per 1000 kelahiran. Artinya Indonesia belum bisa mencapai target yang ditetapkan MDGs. Indonesia masih harus terus berbenah untuk menanggulangi permasalahan kesehatan untuk mencapai target MDGs.

Sebelum dilakukan analisis LORENS, perlu ditentukan dahulu variabel respon dan variabel prediktor. Selanjutnya lakukan analisis statistika deskriptif untuk melihat karakteristik data kematian bayi. Setelah itu, bagi data menjadi dua yaitu data *training* dan data *testing* menggunakan beberapa komparasi.

Proses pengklasifikasian menggunakan metode LORENS dilakukan terlebih dahulu pendekatan algoritma LR CERP. Sama halnya dengan LR CERP, LORENS mempartisi data kedalam  $k$  subruang yang ditentukan secara acak dan distribusi yang sama. Sehingga dapat diasumsikan bahwa pemilihan variabel tidak mengalami bias di setiap subruang. Masing-masing subruang digunakan model LR tanpa pemilihan variabel. Penelitian ini menggunakan 13 faktor yang mempengaruhi angka kematian bayi, di mana faktor tersebut digunakan sebagai variabel prediktor. Faktor yang digunakan adalah berat badan bayi lahir ( $x_1$ ), usia ibu saat melahirkan ( $x_2$ ), Pendidikan ibu ( $x_3$ ), pekerjaan ibu ( $x_4$ ), interval kelahiran ( $x_5$ ), Frekuensi pemeriksaan kehamilan ( $x_6$ ), tempat persalinan ( $x_7$ ), jenis kelamin bayi ( $x_8$ ), pemberian ASI ( $x_9$ ), jenis kelahiran ( $x_{10}$ ), urutan kelahiran ( $x_{11}$ ), tempat tinggal ( $x_{12}$ ), indeks kekayaan keluarga ( $x_{13}$ ). Sedangkan variabel respon dalam penelitian ini adalah kematian bayi ( $y$ ).

Setelah dilakukan pengklasifikasian, akan diperoleh nilai prediksi. Setelah itu lakukan pengamatan terhadap data testing untuk menentukan pengamatan positif atau negatif. Selanjutnya bandingkan kelas *actual* dan prediksi hasil klasifikasi, perbandingan akan berupa kelompok *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Berdasarkan langkah sebelumnya diperoleh hasil ketepatan model berupa *sensitivity*, *specificity*, dan *accuracy*.

Selanjutnya lakukan evaluasi performa model membagi data menjadi  $k$  *folds* atau partisi yang memiliki jumlah seimbang. Berdasarkan partisi tersebut, *cross validation* juga membagi data menjadi data *testing* dan data *training*. Semua partisi digunakan sebagai data *testing* sekaligus data *training*. Penggunaan data sebagai data *testing* atau data *training* dilakukan saat gilirannya masing-masing. Giliran masing-masing data dapat dilakukan karena prosedur berulang dari *cross validation*. Misalkan dalam *cross validation* digunakan sebagai data *testing* sebanyak satu partisi,  $k - 1$  sisanya digunakan sebagai data *training*. Selanjutnya, dilakukan secara berulang hingga semua partisi telah dijadikan sebagai data *testing*. Prosedur ini disebut dengan *k fold cross validation*.

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Pendekatan Penelitian**

Penelitian ini menggunakan pendekatan deskriptif kuantitatif serta studi literatur. Pendekatan deskriptif kuantitatif merupakan proses analisis terkait suatu sampel dan populasi tertentu menggunakan ilmu statistika yang bersifat kuantitatif. Sedangkan studi literatur merupakan pendekatan yang dilakukan dengan mempelajari dan mengkaji teori dari berbagai literatur berupa jurnal, buku, dan sebagainya sebagai referensi yang berkaitan serta dibutuhkan dalam suatu penelitian.

#### **3.2 Jenis dan Sumber Data**

Data yang digunakan pada penelitian ini adalah data sekunder berupa angka kematian bayi (AKB) di Indonesia. Data diperoleh dari data dokumenter hasil sensus kesehatan Badan Pusat Statistik (BPS) tahun 2017.

#### **3.3 Variabel Penelitian**

Variabel yang digunakan dalam penelitian adalah variabel  $x$  dan  $y$ . Variabel  $x$  adalah variabel prediktor berupa faktor-faktor yang mempengaruhi angka kematian bayi. Sedangkan  $y$  adalah variabel respon berupa data kategorik kematian bayi. Tabel 3.1 merupakan variabel yang digunakan dalam penelitian ini serta data lebih jelas dapat di lihat pada Lampiran 1.

Tabel 3.1 Variabel Angka Kematian Bayi

No	Variabel		Jenis	Skala	Keterangan
1	y	Kematian bayi	Bayi Hidup	Nominal	0
			Kematian bayi		1
2	x <sub>1</sub>	Berat Badan Bayi Lahir	BBL Normal	Ordinal	0
			BBL Lebih		1
			BBL Rendah		2
3	x <sub>2</sub>	Usia ibu saat melahirkan	Numerik	Rasio	-
4	x <sub>3</sub>	Pendidikan Ibu	Tamatan ≥ SMA	Ordinal	0
			Tamatan SMP		1
			Tamatan ≤ SD		2
5	x <sub>4</sub>	Pekerjaan Ibu	Bekerja	Nominal	0
			Tidak Bekerja		1
6	x <sub>5</sub>	Interval Kelahiran	>47 bulan	Ordinal	0
			36-47 bulan		1
			24-35 bulan		2
			18-23 bulan		3
			<18 bulan		4
7	x <sub>6</sub>	Frekuensi <i>Antenatal Care</i> (Pemeriksaan Kehamilan)	Sesuai Standar	Nominal	0
			Tidak sesuai standar		1
			Tidak ANC		2
8	x <sub>7</sub>	Tempat Persalinan	Faskes	Nominal	0
			Non Faskes		1
9	x <sub>8</sub>	Jenis Kelamin Bayi	Perempuan	Nominal	0
			Laki-laki		1
10	x <sub>9</sub>	Pemberian ASI	Memberi ASI	Nominal	0
			Tidak memberi ASI		1
11	x <sub>10</sub>	Jenis Kelahiran	Tunggal	Nominal	0
			Kembar		1
12	x <sub>11</sub>	Urutan Kelahiran	Numerik	Rasio	-
13	x <sub>12</sub>	Tempat tinggal	Perkotaan	Nominal	0
			Perdesaan		1
14	x <sub>13</sub>	Indeks Kekayaan Keluarga	Sangat kaya	Ordinal	0
			Kaya		1
			Menengah		2
			Miskin		3
			Sangat miskin		4

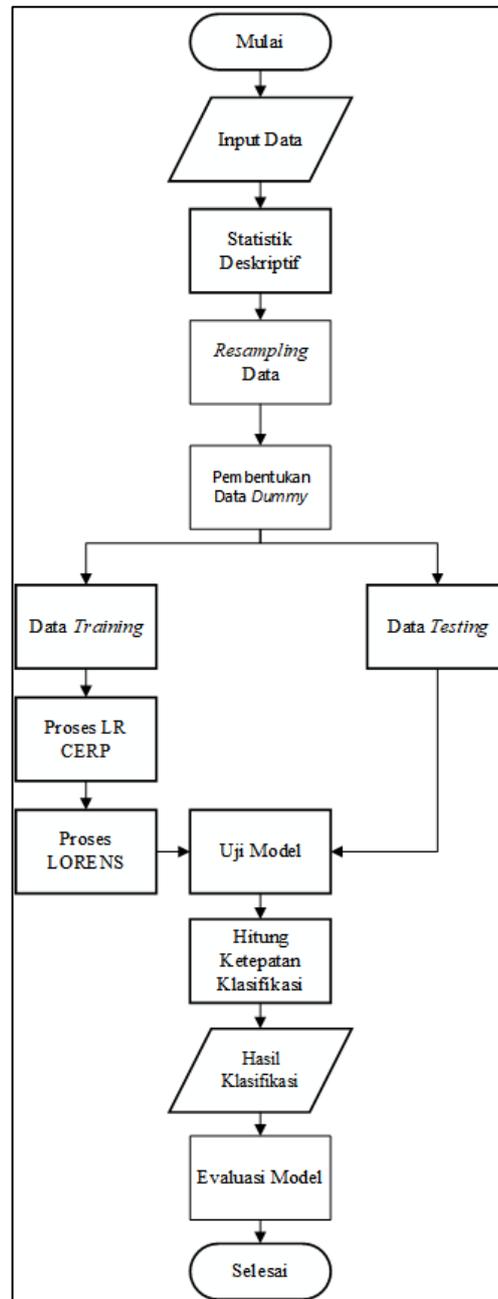
### 3.4 Tahapan Penelitian

Proses dalam mengklasifikasi angka kematian bayi menggunakan metode *logistic regression ensemble* (LORENS) dilakukan dengan tahapan sebagai berikut:

1. Melakukan statistik deskriptif untuk mengetahui karakteristik data;
2. Melakukan *resampling* berupa *over sampling* untuk mengatasi data *imbalance*;
3. Bentuk data kategorik dalam bentuk *dummy*;
4. Membagi data menjadi data *training* dan data *testing*;
5. Menentukan jumlah  $l$  partisi dan  $m$  *ensemble* yang akan diterapkan;
6. Partisi variabel prediktor kedalam  $k$  subruang;
7. Membangun model LR di setiap partisi dari data *training*;
8. Menghitung nilai prediksi setiap partisi dengan mensubstitusikan data *testing* ke model;
9. Setelah nilai prediksi diperoleh, hitung rata-ratanya;
10. Lakukan langkah (4) hingga (8) sampai terbentuk  $n$  *ensemble*;
11. Mencari nilai prediksi terbanyak dari pengamatan pada semua *ensemble*;
12. Menghitung nilai *threshold* paling optimal;
13. Klasifikasi setiap nilai prediksi menjadi 0 atau 1 berdasarkan nilai *threshold*;
14. Bandingkan nilai prediksi dengan data yang dimiliki, lalu kelompokkan ke dalam kelompok TP, TN, FP, dan FN;
15. Menghitung ketepatan hasil klasifikasi dengan menghitung nilai *sensitivity*, *specificity*, dan *accuracy*;

16. Lakukan evaluasi performa model menggunakan *cross validation*; dan
17. Menarik kesimpulan dari hasil klasifikasi.

### 3.5 Diagram Alir



Gambar 3.1 Diagram Alir Penelitian

## BAB IV

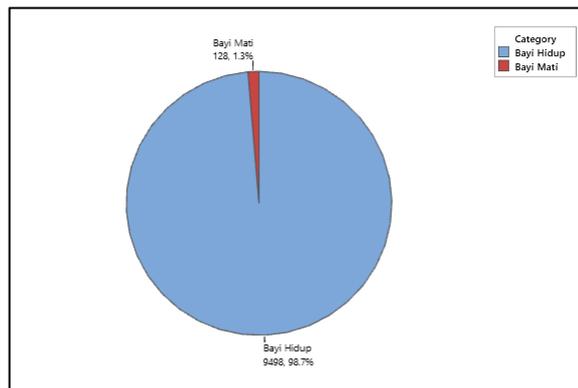
### HASIL DAN PEMBAHASAN

#### 4.1 Statistik Deskriptif

Penelitian ini menggunakan data kematian bayi di Indonesia tahun 2017 yang terdiri dari 9626 data. Statistik deskriptif dalam penelitian ini berupa perbandingan data pada variabel kategorik, namun pada data bertipe numerik akan di cari karakteristik data berupa *mean*, *median*, standar deviasi, minimum, dan maksimum. Data dalam penelitian ini terdiri dari satu variabel respon berupa variabel kematian bayi ( $y$ ) dan 13 variabel prediktor berupa berat badan bayi lahir ( $x_1$ ), usia ibu saat melahirkan ( $x_2$ ), pendidikan ibu ( $x_3$ ), pekerjaan ibu ( $x_4$ ), interval kelahiran ( $x_5$ ), frekuensi pemeriksaan kehamilan ( $x_6$ ), tempat persalinan ( $x_7$ ), jenis kelamin bayi ( $x_8$ ), pemberian ASI ( $x_9$ ), jenis kelahiran ( $x_{10}$ ), urutan kelahiran ( $x_{11}$ ), tempat tinggal ( $x_{12}$ ), indeks kekayaan keluarga ( $x_{13}$ ).

##### 4.1.1 Statistik Deskriptif Variabel Respon

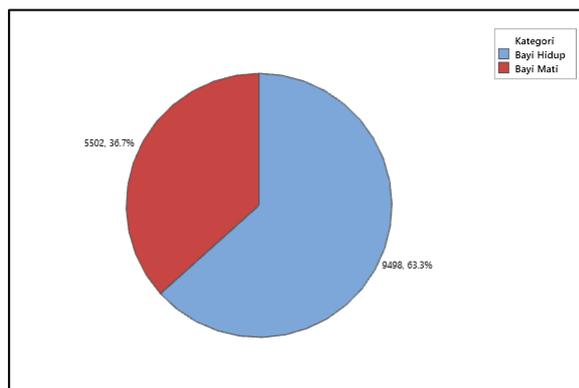
Variabel ( $y$ ) atau kematian bayi yang berupa data kategorik bayi mati dan bayi hidup memiliki perbandingan data yang tidak seimbang atau *imbalance*. Berikut merupakan perbandingan data bayi hidup dan bayi mati ketika *imbalance*.



Gambar 4.1 *Imbalance Data* pada Sensus Kematian Bayi

Berdasarkan Gambar 4.1 diketahui bahwa 128 bayi dari 9626 kelahiran bayi atau sebanyak 1,3% kasus mengalami kematian. Sisanya yaitu 9498 bayi dari 9626 kelahiran bayi atau sebanyak 98,7% kasus bayi hidup. Hal ini menunjukkan bahwa terjadi 14 kematian dari 1000 kelahiran bayi.

Diantara data bayi hidup dan data bayi mati bersifat *imbalance* atau tidak seimbang. Apabila dilakukan proses analisis data terhadap data yang *imbalance* mengakibatkan kekeliruan analisis dan sulit untuk menghitung *sensitivity*, *specificity*, serta menghasilkan model yang sukar dipahami. Sehingga perlu dilakukan *resampling* untuk mengatasinya dengan memperbanyak data bayi mati. *Resampling* yang digunakan dalam penelitian ini adalah *over sampling*. *Over sampling* digunakan karena konsep penyeimbangan data yang dilakukan *over sampling* tidak perlu mengeliminasi data satupun, tetapi memperbanyak data minoritas berdasarkan kesamaan sifatnya. Berdasarkan *over sampling* yang dilakukan pada data maka data yang sebelumnya 9626 menjadi sebanyak 15000 data. Berikut merupakan perbandingan data bayi hidup dan bayi mati setelah dilakukan *over sampling*.



Gambar 4.2 *Balance Data* pada Sensus Kematian Bayi

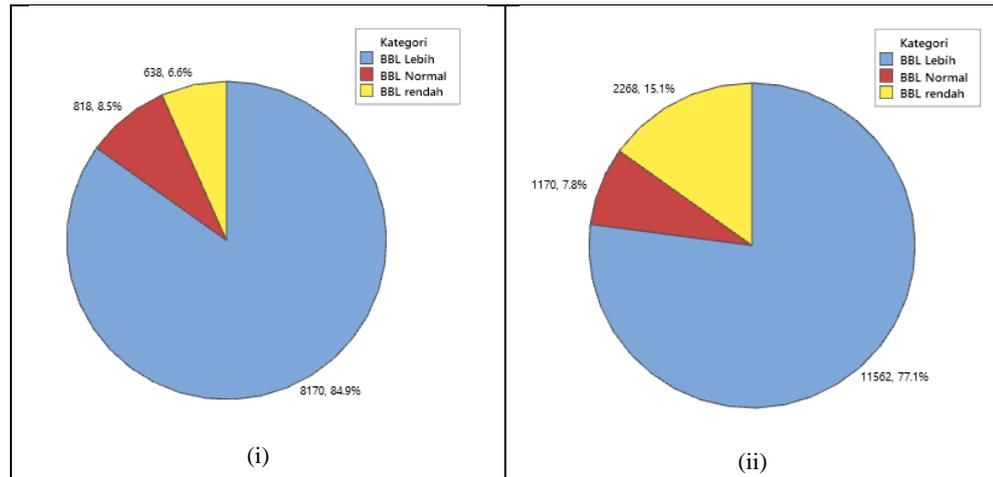
Setelah dilakukan *over sampling* pada data, maka data bayi mati dapat ditingkatkan. Antara kategori bayi mati dan bayi hidup telah seimbang (*balance*) dan tidak terjadi ketimpangan yang berarti. Perbandingan data bayi adalah 63,3% atau 9495 data terdiri dari bayi hidup dan sisanya 36,7% atau 5505 data adalah bayi mati. Data ini memberikan penambahan sebanyak 43 kali terhadap data bayi mati yang semula 128 kematian bayi menjadi 5505 kematian bayi.

#### 4.1.2 Statistik Deskriptif Variabel Prediktor

Statistik deskriptif pada variabel prediktor yang terdiri dari 13 variabel penyebab kematian bayi memiliki karakteristik sebagai berikut.

1. Variabel Berat Badan Bayi Lahir ( $x_1$ )

Variabel Berat Badan Bayi Lahir ( $x_1$ ) berupa data kategorik dengan 3 kategori yaitu berat badan bayi lahir lebih, berat badan bayi lahir kurang, dan berat badan bayi lahir normal. Berikut adalah grafik perbandingan ketiga kategori ketika *imbalance data* dan *balance data*.



Gambar 4.3 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Berat Badan Bayi Lahir

Gambar 4.3 menunjukkan baik ketika *imbalance data* maupun *balance data* terdapat perubahan pada karakteristik data. Data kategori BBL lebih menunjukkan mayoritas bayi di Indonesia lahir dengan keadaan berat badan bayi lahir lebih. Pada *balance data* menunjukkan perubahan berupa penambahan presentase pada BBL rendah, dan pengurangan presentase pada BBL normal dan BBL lebih.

## 2. Variabel Usia Ibu Saat Melahirkan ( $x_2$ )

Variabel Usia Ibu Saat Melahirkan ( $x_2$ ) berupa data numerik yaitu usia ibu saat melahirkan. Berikut adalah grafik perbandingan usia ibu saat melahirkan ketika *imbalance data* dan *balance data*.

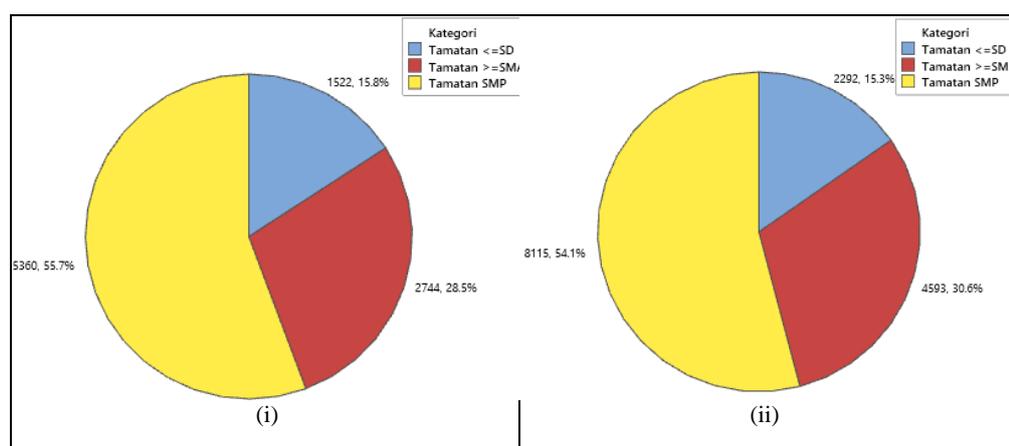
Tabel 4.1 *Imbalance Data* dan *Balance Data* pada Sensus Ibu Saat Melahirkan

<i>Imbalance data</i>		<i>Balance data</i>	
Minimum	15	Minimum	15
Sdt. Dev.	5,437	Sdt. Dev.	5,517
Rata-rata	31,507	Rata-rata	32,117
Median	31	Median	32
Maksimum	48	Maksimum	48

Tabel 4.1 menunjukkan baik ketika *imbalance data* maupun *balance data* mengalami perubahan karakteristik. Perubahan pada standar deviasi sebelumnya 5,437 pada *imbalance data* menjadi 5,517 pada *balance data*. Selain itu rata-rata data juga mengalami sedikit penambahan yaitu 31,507 pada *imbalance data* menjadi 32,117 pada *balance data*.

### 3. Variabel Pendidikan Ibu ( $x_3$ )

Variabel Pendidikan Ibu ( $x_3$ ) merupakan variabel bersifat kategorik yang terdiri dari tamatan  $\leq$  SD, tamatan SMP, dan tamatan  $\geq$  SMA. Berikut adalah grafik perbandingan ketiga kategori ketika *imbalance data* dan *balance data*.



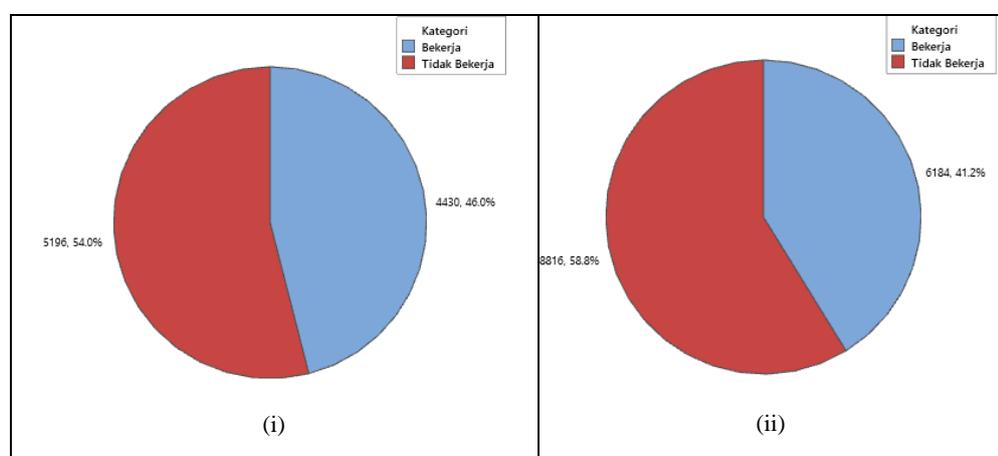
Gambar 4.4 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Pendidikan Ibu

Gambar 4.4 menunjukkan baik ketika *imbalance data* maupun *balance data* menunjukkan sedikit perubahan karakteristik. Perubahan ini pengurangan presentase pada lulusan  $\leq$  SD dan lulusan  $\geq$  SMA, serta penambahan pada lulusan SMP. Saat ini mayoritas ibu adalah lulusan SMP sederajat. Hal ini menunjukkan bahwa program pemerintah berupa wajib belajar 9 tahun yang disahkan pada tahun 1995 sudah terlaksana. Namun

sejak tahun 2015 wajib belajar di Indonesia diberlakukan wajib belajar 12 tahun, yaitu minimal sampai SMA sederajat.

#### 4. Variabel Pekerjaan Ibu ( $x_4$ )

Variabel Pekerjaan Ibu ( $x_4$ ) merupakan variabel bersifat kategorik yang terdiri dari tamatan bekerja dan tidak bekerja. Berikut adalah grafik perbandingan kedua kategori ketika *imbalance data* dan *balance data*.



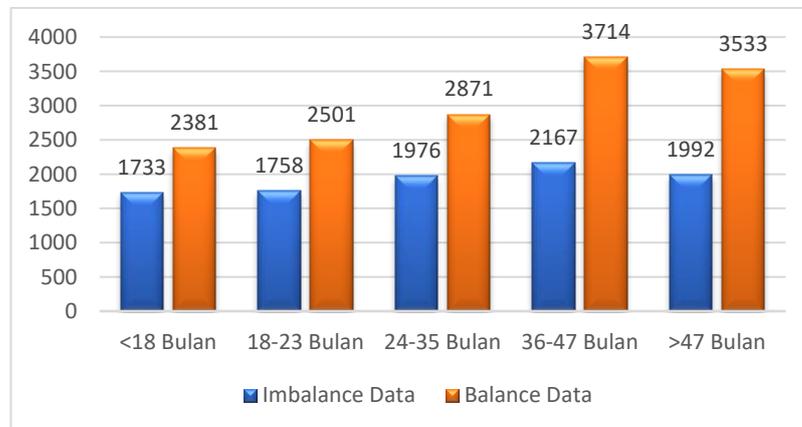
Gambar 4.5 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Pekerjaan Ibu

Gambar 4.5 menunjukkan baik ketika *imbalance data* maupun *balance data* terdapat perubahan karakteristik sebesar 4,8%. Perubahan ini berupa pengurangan pada kategori ibu bekerja sekaligus penambahan pada kategori ibu tidak bekerja sebesar 54% pada *imbalance data* menjadi 58,8% pada *balance data*.

#### 5. Variabel Interval Kelahiran ( $x_5$ )

Variabel Interval Kelahiran ( $x_5$ ) merupakan variabel bersifat kategorik yang terdiri dari >47 bulan, 36-47 bulan, 24-35 bulan, 18-23 bulan, dan <18

bulan. Berikut adalah grafik perbandingan kelima kategori ketika *imbalance data* dan *balance data*.

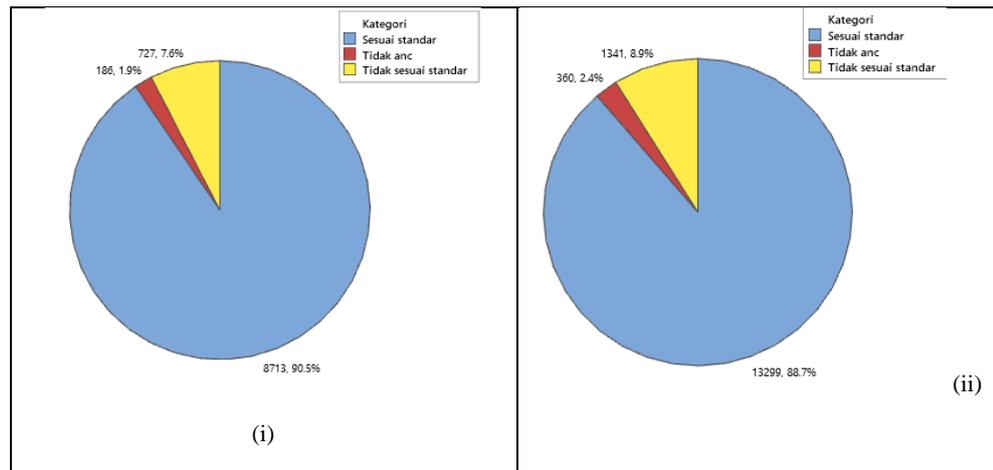


Gambar 4.6 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Interval Kelahiran

Gambar 4.6 menunjukkan bahwa ketika *imbalance data* maupun *balance data* terdapat perubahan pada semua kategori. Perubahan yang terjadi pada *balance data* didasarkan pada *imbalance data*. Dari kelima kategori, bayi berusia 36-47 bulan merupakan kategori dengan jumlah terbanyak. Sedangkan bayi berusia <18 bulan merupakan kategori dengan jumlah tersedikit.

#### 6. Variabel Frekuensi *Antenatal Care* (anc) atau Pemeriksaan Kehamilan ( $x_6$ )

Variabel Frekuensi *Antenatal Care* (anc) atau Pemeriksaan Kehamilan ( $x_6$ ) merupakan variabel kategorik yang memiliki 3 kategori antara lain pemeriksaan kehamilan sesuai standar, tidak melakukan pemeriksaan kehamilan, dan pemeriksaan kehamilan tidak sesuai standar. Berikut adalah grafik perbandingan ketiga kategori ketika *imbalance data* dan *balance data*.



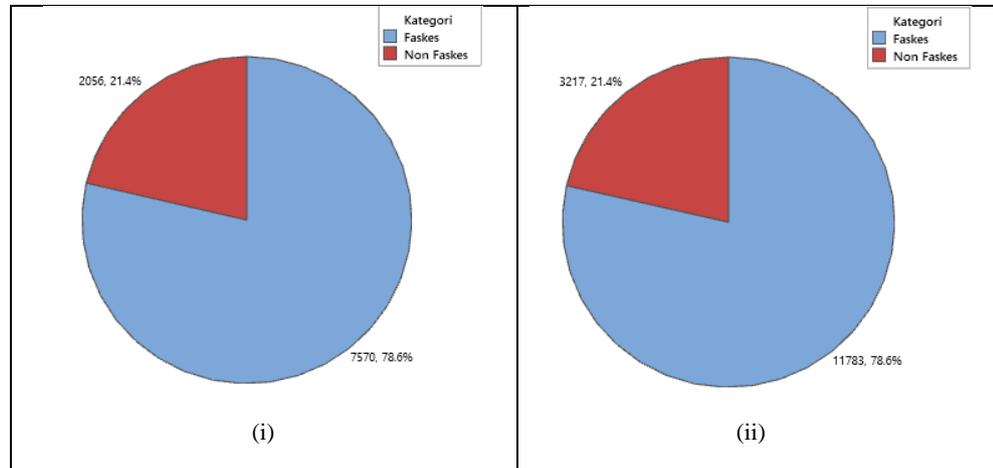
Gambar 4.7 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Frekuensi pemeriksaan kehamilan

Gambar 4.7 menunjukkan bahwa dari ketiga kategori pemeriksaan kehamilan, baik ketika *imbalance data* maupun *balance data* terdapat perubahan berupa penambahan presentase pada kategori tidak melakukan pemeriksaan kehamilan dan kategori tidak sesuai standar. Sedangkan pada kategori pemeriksaan sesuai standar mengalami sedikit penurunan. Secara umum kepedulian ibu dalam melakukan pemeriksaan kehamilan sudah sangat baik, hanya sebagian kecil ibu yang perlu diberi edukasi terkait pentingnya pemeriksaan kehamilan.

#### 7. Variabel Tempat Persalinan ( $x_7$ )

Variabel Tempat Persalinan ( $x_7$ ) merupakan variabel kategorik yang memiliki 2 kategori antara lain Faskes dan Non Faskes. Faskes artinya Ibu melahirkan di tempat dengan fasilitas kesehatan. Sedangkan non faskes artinya ibu melahirkan di tempat tanpa fasilitas kesehatan seperti dukun melahirkan dan sebagainya. Banyak faktor yang menjadi penyebab ibu tidak melahirkan ditempat dengan fasilitas kesehatan, antara lain karena lokasinya

yang masih di pedalaman, kurang meratanya fasilitas kesehatan, dan lain-lain. Berikut adalah grafik perbandingan kedua kategori ketika *imbalance data* dan *balance data*.

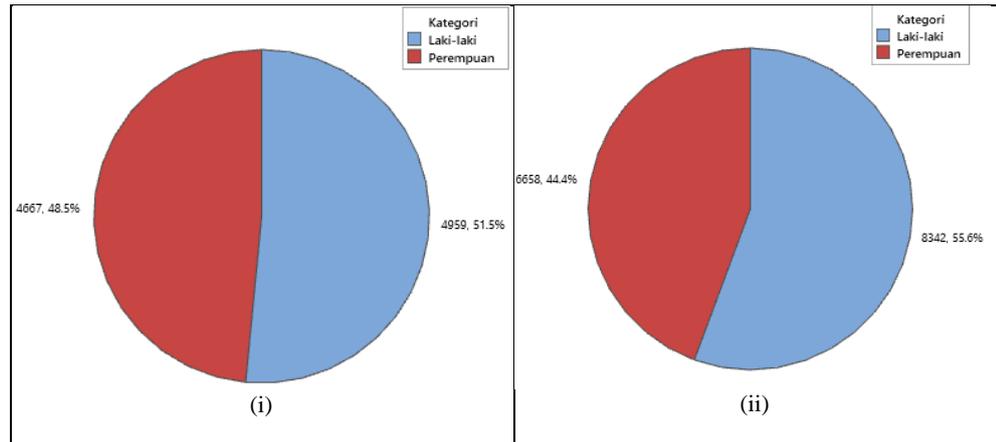


Gambar 4.8 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Tempat Persalinan

Gambar 4.8 menunjukkan bahwa dari kedua kategori tempat persalinan ibu saat melahirkan baik ketika *imbalance data* maupun *balance data* tidak mengalami perubahan karakteristik. Hasil menunjukkan kesamaan sebesar 78,6% tempat persalinan dengan fasilitas kesehatan sebelum maupun sesudah *over sampling*. Hal ini menunjukkan mayoritas masyarakat Indonesia sudah menggunakan fasilitas kesehatan saat melahirkan. Sedangkan sisanya 21,4% ibu melakukan persalinan di tempat tanpa fasilitas kesehatan.

#### 8. Variabel Jenis Kelamin Bayi ( $x_8$ )

Variabel Jenis Kelamin Bayi ( $x_8$ ) merupakan variabel kategorik yang memiliki 2 kategori antara lain perempuan dan laki-laki. Berikut adalah grafik perbandingan kedua kategori ketika *imbalance data* dan *balance data*.

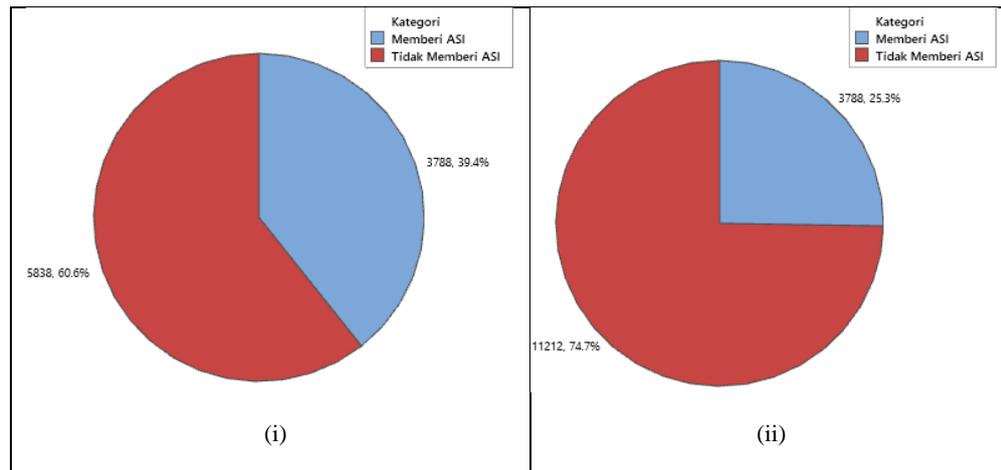


Gambar 4.9 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Jenis Kelamin Bayi

Gambar 4.9 menunjukkan bahwa dari kedua Jenis kelamin bayi baik ketika *imbalance data* maupun *balance data* mengalami sedikit perubahan. Perubahan ini berupa pengurangan kategori perempuan sekaligus penambahan kategori laki-laki. Pada jenis kelamin laki-laki mengalami peningkatan presentase sebesar 4,1%, yang sebelumnya adalah 51,5% menjadi 55,6%. Sampel ini menunjukkan bahwa perbandingan antara bayi perempuan dengan bayi laki-laki memiliki jumlah yang tidak berbeda jauh.

#### 9. Variabel Pemberian ASI ( $x_9$ )

Variabel Pemberian ASI ( $x_9$ ) merupakan variabel kategorik yang memiliki 2 kategori antara lain memberi ASI dan tidak memberi ASI. Gambar 4.10 menunjukkan bahwa antara pemberian ASI ketika *imbalance data* maupun *balance data* mengalami perubahan yang cukup signifikan. Kategori tidak memberikan ASI pada *Imbalance data* menunjukkan 60,6%, bertambah menjadi 74,7% saat *balanced data*

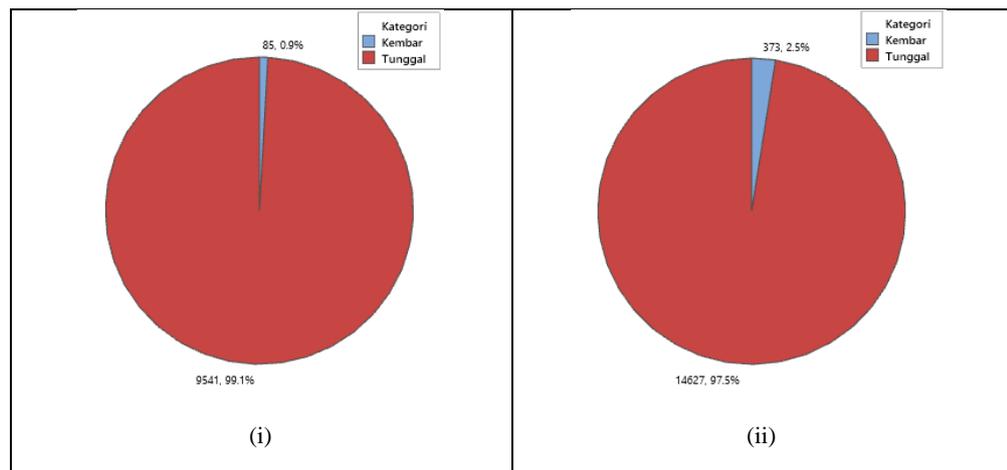


Gambar 4.10 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Pemberian ASI

Perbandingan antara kategori ibu memberi ASI dan ibu tidak memberi ASI, mayoritas ibu tidak memberikan ASI kepada bayinya. Hal ini sangat disayangkan sekali bahwa sangat banyak ibu yang tidak memberikan asi kepada bayinya. Karena kebutuhan bayi akan ASI masih sangatlah tinggi terutama saat bayi berusia 0 s.d 24 bulan.

#### 10. Variabel Jenis Kelahiran ( $x_{10}$ )

Variabel Jenis Kelahiran ( $x_{10}$ ) merupakan variabel kategorik yang memiliki 2 kategori antara lain memberi kelahiran tunggal dan kelahiran kembar. Gambar 4.11 menunjukkan ketika *imbalance data* maupun *balance data* mengalami perubahan berupa penambahan presentase pada kategori bayi lahir kembar sebesar 1,6%. Serta pengurangan presentase pada kategori bayi lahir tunggal. Kedua data menunjukkan mayoritas bayi lahir dengan kelahiran tunggal dan jarang sekali terjadi kelahiran kembar Hal ini dikarenakan kelahiran kembar dipengaruhi oleh gen dari orang tua.



Gambar 4.11 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Jenis Kelahiran

### 11. Variabel Urutan Kelahiran ( $x_{11}$ )

Variabel Urutan Kelahiran ( $x_{11}$ ) berupa data numerik yaitu urutan kelahiran bayi. Berikut adalah grafik perbandingan urutan kelahiran ketika *imbalance data* dan *balance data*.

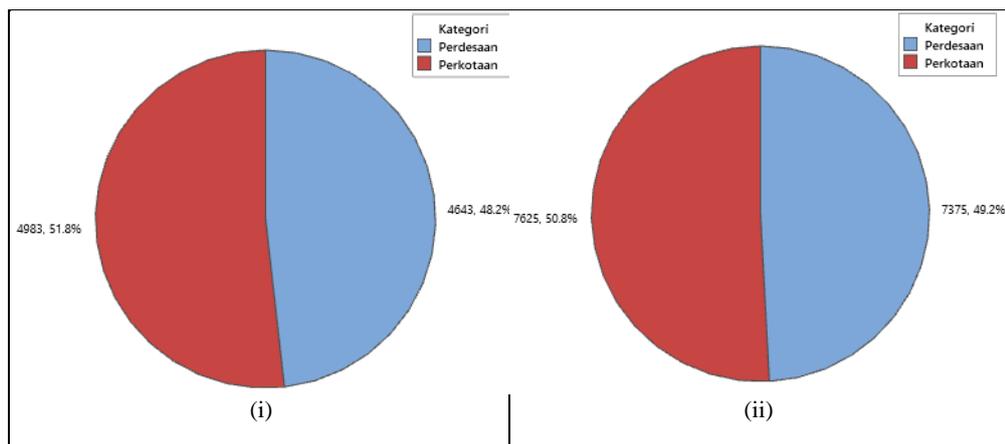
Tabel 4.2 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Urutan Kelahiran

<i>Imbalance data</i>		<i>Balance data</i>	
Minimum	2	Minimum	2
Sdt. Dev.	1,93	Sdt. Dev.	1,27
Rata-rata	2,93	Rata-rata	2,93
Median	3	Median	3
Maksimum	12	Maksimum	12

Tabel 4.2 menunjukkan bahwa urutan kelahiran bayi, data minimum adalah anak urutan kedua. Tidak terjadi perubahan pada variabel urutan kecuali pada standar deviasinya. Sedangkan maksimum urutan bayi adalah pada urutan anak ke-12. Rata-rata urutan kelahiran dalam penelitian ini adalah urutan anak ketiga dengan standar deviasi adalah 1,93 pada *imbalance data*, dan 1,27 pada *balance data*.

## 12. Variabel Tempat Tinggal ( $x_{12}$ )

Variabel Tempat Tinggal ( $x_{12}$ ) merupakan variabel kategorik yang memiliki 2 kategori antara lain perdesaan dan perkotaan. Grafik perbandingan kedua kategori sebelum dan sesudah *over sampling* adalah sebagai berikut.

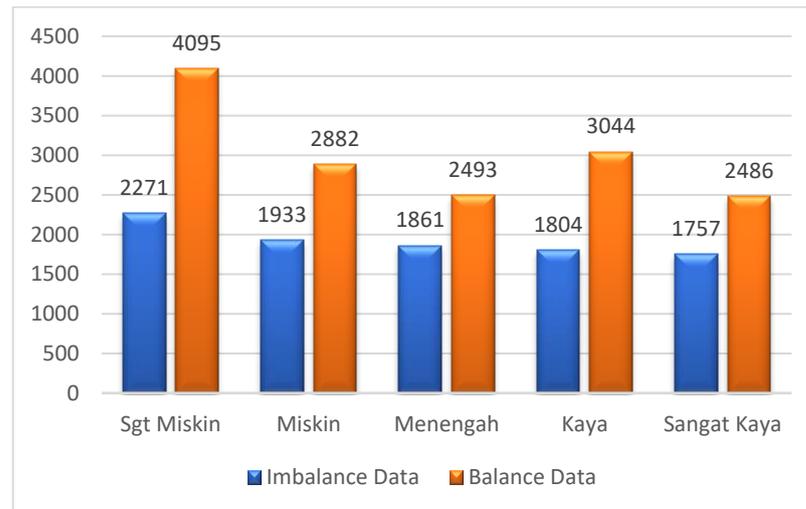


Gambar 4.12 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Tempat Tinggal

Gambar 4.12 menunjukkan bahwa dari kedua kategori tempat tinggal, terdapat sedikit perubahan karakteristik. Perubahan tersebut berupa penambahan kategori perkotaan sekaligus pengurangan kategori perdesaan sebesar 1%. Baik ketika *imbalance data* dan *balance data* perbandingan antara ibu yang bertempat tinggal di perdesaan maupun perkotaan memiliki jumlah yang hampir sama.

## 13. Variabel Indeks Kekayaan Keluarga ( $x_{13}$ )

Variabel indeks kekayaan keluarga memiliki lima kategori interval kelahiran yang terdiri dari sangat miskin, miskin, menengah, kaya, dan sangat kaya. Berikut adalah grafik perbandingan kelima kategori ketika *imbalance data* dan *balance data*.



Gambar 4.13 *Imbalance Data* (i) dan *Balance Data* (ii) pada Sensus Indeks Kekayaan Keluarga

Gambar 4.13 menunjukkan ketika *imbalance data* terjadi penurunan jumlah saat kategori sangat miskin menuju kategori sangat kaya. Artinya keluarga dengan indeks semakin kaya, maka jumlahnya akan semakin sedikit. Sedangkan pada *balance data* mayoritas indeks kekayaan keluarga adalah sangat miskin, namun tidak terjadi penurunan seperti halnya pada *imbalance data*, terutama pada indeks kekayaan keluarga kategori kaya mengalami peningkatan yang cukup tinggi dibandingkan kategori lain.

## 4.2 Pembentukan Data *Dummy*

Sebelum dilakukannya analisis *Logistic Regression Ensemble* (LORENS), perlu dilakukan perubahan data kategorik dalam bentuk *dummy*. Hal ini dikarenakan data dengan skala kategorik dalam penelitian ini memiliki kategori yang berbeda-beda. Berikut merupakan variabel prediktor setelah dilakukan *dummy*.

Tabel 4.3 Variabel Prediktor Hasil *Dummy*

$x_1D_1$	$x_1D_2$	$x_2$	$x_3D_1$	$x_3D_2$	$x_4D_1$	$x_5D_1$	$x_5D_2$	$x_5D_3$	$x_5D_4$	$x_6D_1$	$x_6D_2$
0	1	38	0	1	1	1	0	0	0	0	0
1	0	28	0	1	0	0	1	0	0	0	0
1	0	39	0	1	0	0	0	1	0	0	0
0	0	35	0	1	0	0	1	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	0	31	0	1	1	1	0	0	0	0	0

Tabel 4.4 Lanjutan Variabel Prediktor Hasil *Dummy*

$x_7D_1$	$x_8D_1$	$x_9D_1$	$x_{10}D_1$	$x_{11}$	$x_{12}D_1$	$x_{13}D_1$	$x_{13}D_2$	$x_{13}D_3$	$x_{13}D_4$
0	1	1	1	6	1	0	0	0	1
0	1	1	0	4	1	0	1	0	0
0	1	1	0	5	1	0	0	0	1
1	1	0	0	8	1	0	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	1	0	4	0	0	0	1	0

Tabel 4.3 dan 4.4 menunjukkan hasil *dummy* yang dibentuk berdasarkan variabel prediktor. Sebelum dilakukan proses *dummy* variabel prediktor penelitian adalah 13, setelah dilakukan *dummy* variabel prediktor bertambah menjadi 22 sesuai kategorik masing-masing variabel. Variabel  $x_1D_1$  menunjukkan variabel berat badan bayi lahir dengan *dummy* pertama, di mana *dummy* pertama adalah berat badan bayi lahir lebih dengan kontrol berat badan bayi lahir normal. Selanjutnya variabel  $x_1D_2$  menunjukkan variabel berat badan bayi lahir dengan *dummy* kedua, di mana *dummy* kedua adalah berat badan bayi lahir kurang. Hal serupa diaplikasikan hingga semua data kategorik pada  $x_{13}$  atau indeks kekayaan keluarga yang terdiri dari empat variabel *dummy*.

### 4.3 Analisis LORENS

#### 4.3.1 Pembagian data *Training* dan *Testing*

Proses analisis pada *Logistic Regression Ensemble* (LORENS) dilakukan dengan membagi data menjadi data *training* dan data *testing*. Data *training*

digunakan sebagai data latih, sedangkan data *testing* digunakan sebagai data uji. Umumnya pembagian data ini dibagi menjadi 70%:30%, 75%:25%,..., 90%:10%. Pembagian data ini dilakukan untuk memperoleh perbandingan antara data *training* dan data *testing* dengan nilai akurasi terbaik. Pada penelitian ini dilakukan uji coba pada masing-masing perbandingan data *training* dan data *testing* dengan hasil sebagai berikut.

Tabel 4.5 Kombinasi Data *Training* dan Data *Testing*

<b>Perbandingan data <i>training</i> dan data <i>testing</i></b>	<b>Akurasi(%)</b>
70% : 30%	80,11
75% : 25%	77,89
80% : 20%	79,67
85% : 15%	80,93
90% : 10%	81,47

Tabel 4.5 menunjukkan bahwa dari beberapa kombinasi perbandingan data *training* dan *testing* memiliki akurasi yang cukup baik. Namun dari perbandingan antara data *training* dan *testing* dengan akurasi terbaik adalah menggunakan perbandingan 90% : 10% dengan nilai akurasi 81,47%. Oleh karenanya, penelitian ini menggunakan perbandingan 90% : 10% sebagai pembagian data *training* dan data *testing*.

#### **4.3.2 Partisi dan *Ensemble***

Penelitian ini dilakukan dengan membagi data dalam beberapa partisi secara acak. Berdasarkan 2.3.1 untuk memperoleh partisi optimal dapat dilakukan berdasarkan jumlah data dan jumlah variabel. Banyak data yang digunakan dalam penelitian ini terdiri dari 15000 data. Sedangkan banyak variabel prediktor data terdiri dari 22 variabel. Sehingga pada penelitian ini digunakan persamaan 2.13 untuk memperoleh nilai partisi yang optimal.

$$K = \frac{p}{q} = \left\lfloor \frac{15000}{7500} \right\rfloor = 2$$

$$K = \frac{p}{q} = \left\lfloor \frac{15000}{5000} \right\rfloor = 3$$

⋮

$$K = \frac{p}{q} = \left\lfloor \frac{15000}{682} \right\rfloor = 22$$

Berdasarkan partisi optimal yang diperoleh menggunakan persamaan 2.13 di mana  $q$  adalah bilangan integer kurang dari  $n$  menghasilkan partisi yaitu 2,3, ... ,22. Berikut merupakan hasil akurasi berdasarkan masing-masing partisi.

Tabel 4.6 Kombinasi Partisi dan *Ensemble*

<i>Ensemble</i>	Partisi	Akurasi (%)	<i>Ensemble</i>	Partisi	Akurasi (%)
5	2	79,47	10	2	80,13
	3	77,4		3	77,47
	4	76,87		4	77
	5	75,13		5	76,07
	6	74,4		6	75,33
	⋮	⋮		⋮	⋮
	22	63,6		22	63,6
15	2	81,2	20	<b>2</b>	<b>81,47</b>
	3	78,26		3	78,6
	4	77,33		4	76,8
	5	76,47		5	76,87
	6	75,2		6	74,87
	⋮	⋮		⋮	⋮
	22	63,6		22	63,6
25	2	81,13	30	2	80,8
	3	78,13		3	78,4
	4	76,87		4	76,93
	5	76,53		5	76,87
	6	74,33		6	74,4
	⋮	⋮		⋮	⋮
	22	63,6		22	63,6

Berdasarkan Tabel 4.6 diperoleh bahwa diantara partisi-partisi menghasilkan nilai akurasi paling optimal yaitu 81,47%. Nilai *ensemble* yang

digunakan dalam penelitian ini adalah 20 *ensemble* dimana setiap *ensemble* akan menghasilkan dua model sesuai partisinya. Pada *ensemble* 20 yang digunakan, maka banyak model LR yang terbentuk pada 2 partisi adalah sebanyak 40 model.

Setelah partisi dan *ensemble* diperoleh maka dapat variabel-variabel prediktor dapat dikelompokkan. Pengelompokkan dilakukan secara acak sesuai nilai partisi dan *ensemble*. Berikut merupakan hasil pengelompokkan partisi variabel prediktor penelitian ini.

Tabel 4.7 Pengelompokkan Variabel berdasarkan Partisi dan *Ensemble*

Variabel	<i>Ensemble</i>						
	1	2	3	4	5	...	20
$x_1D_1$	1	1	1	1	1	...	2
$x_1D_2$	1	2	2	2	1	...	2
$x_2$	2	1	2	2	2	...	1
$x_3D_1$	2	1	1	2	1	...	2
$x_3D_2$	1	2	1	1	1	...	2
$x_4D_1$	2	1	1	2	2	...	1
$x_5D_1$	2	1	2	1	1	...	1
$x_5D_2$	2	2	2	1	2	...	2
$x_5D_3$	2	1	2	2	2	...	2
$x_5D_4$	1	2	2	2	1	...	1
$x_6D_1$	1	2	1	1	1	...	1
$x_6D_2$	2	2	1	1	1	...	1
$x_7D_1$	1	1	2	2	1	...	2
$x_8D_1$	1	2	2	2	2	...	1
$x_9D_1$	1	1	1	2	2	...	2
$x_{10}D_1$	2	2	1	1	2	...	1
$x_{11}$	1	2	1	2	2	...	1
$x_{12}D_1$	1	2	1	1	1	...	2
$x_{13}D_1$	2	1	2	1	1	...	2
$x_{13}D_2$	2	2	2	2	2	...	1
$x_{13}D_3$	1	1	1	1	2	...	2
$x_{13}D_4$	2	1	2	1	2	...	1

Tabel 4.7 menunjukkan bahwa dari 22 variabel prediktor, masing-masing *ensemble* dibagi menjadi dua partisi, di mana tiap *ensemble* menghasilkan pengelompokkan partisi yang berbeda-beda. Pada *ensemble*

pertama, menunjukkan hasil partisi pertama terdiri dari variabel  $x_1D_1, x_1D_2, x_3D_2, x_5D_4, x_{6,1}, x_7D_1, x_8D_1, x_9D_1, x_{11}, x_{12}D_1$ , dan  $x_{13}D_3$ . Sedangkan pada partisi kedua yaitu terdiri dari  $x_2, x_3D_1, x_4D_1, x_5D_1, x_5D_2, x_5D_3, x_6D_2, x_{10}D_1, x_{13}D_1, x_{13}D_2$ , dan  $x_{13}D_4$ . Pada *ensemble* kedua menunjukkan variabel yang masuk dalam partisi pertama adalah antara lain  $x_1D_1, x_2, x_3D_1, x_4D_1, x_5D_1, x_5D_3, x_7D_1, x_9D_1, x_{13}D_1, x_{13}D_3$ , dan  $x_{13}D_4$ . Sedangkan pada partisi kedua menunjukkan variabel antara lain,  $x_1D_2, x_3D_2, x_5D_2, x_5D_4, x_6D_1, x_6D_2, x_8D_1, x_{10}D_1, x_{11}, x_{12}D_1$ , dan  $x_{13}D_2$ .

### 4.3.3 Model LR pada Data Training

Model LR terbentuk pada 20 *ensemble* adalah 20 model LR per partisi. Sehingga dalam penelitian ini model LR yang terbentuk dari 2 partisi adalah sebanyak 40 model LR. Model ini dibangun berdasarkan estimasi parameter pada Persamaan 2.10.

Tabel 4.8 Koefisien Model LR pada Setiap Partisi

<i>Intercept</i>	<i>Ensemble</i>						
	1	2	3	4	5	...	20
Partisi ke-1	-20.55	-20.47	-18.39	-0.19	-0.92	...	-2.58
Partisi ke-2	-2.50	-1.88	-2.93	-21.06	-20.63	...	-18.65

Tabel 4.8 menunjukkan koefisien model LR yang terbentuk pada masing-masing partisi. Sedangkan Tabel 4.9 menunjukkan koefisien model LR yang terbentuk pada masing-masing *ensemble* yang sebelumnya di peroleh. Nilai dari koefisien tersebut menunjukkan pengaruh yang dibangun masing-masing variabel terhadap kematian bayi.

Tabel 4.9 Koefisien Model LR pada Setiap Variabel

Variabel	Ensemble						
	1	2	3	4	5	...	20
$x_1D_1$	0.380	-1.106	-0.929	-1.036	-0.047	...	-0.061
$x_1D_2$	2.003	1.711	1.746	1.789	1.700	...	1.627
$x_2$	0.062	0.064	0.059	0.022	0.046	...	0.038
$x_3D_1$	-0.099	-0.186	-0.103	-0.143	-0.238	...	-0.246
$x_3D_2$	0.122	0.126	0.078	0.206	0.083	...	0.216
$x_4D_1$	-0.523	-0.332	-0.282	-0.486	-0.342	...	-0.551
$x_5D_1$	0.223	1.088	0.085	0.316	0.343	...	0.296
$x_5D_2$	-0.376	-0.413	-0.449	-0.237	-0.928	...	-0.880
$x_5D_3$	-0.437	-0.740	-0.623	-1.302	-1.161	...	-1.111
$x_5D_4$	-1.066	-0.562	-0.673	-1.452	-0.374	...	-0.444
$x_6D_1$	0.275	0.248	0.336	0.409	0.503	...	0.298
$x_6D_2$	0.262	0.156	0.115	0.188	0.481	...	0.217
$x_7D_1$	-0.169	-0.437	0.010	-0.100	-0.038	...	-0.139
$x_8D_1$	0.586	0.632	0.534	0.586	0.581	...	0.614
$x_9D_1$	18.665	18.888	18.382	19.838	18.679	...	18.741
$x_{10}D_1$	2.131	1.160	1.091	1.682	1.723	...	1.717
$x_{11}$	0.304	0.260	0.219	0.249	0.145	...	0.165
$x_{12}D_1$	0.273	0.077	0.156	-0.105	0.212	...	0.277
$x_{13}D_1$	0.432	0.647	0.259	0.524	0.330	...	0.238
$x_{13}D_2$	-0.175	-0.547	-0.415	-0.723	-0.234	...	-0.337
$x_{13}D_3$	-0.123	0.497	-0.151	0.135	0.011	...	-0.038
$x_{13}D_4$	0.599	1.161	0.601	0.595	0.559	...	0.319

Warna	Keterangan
	Partisi 1
	Partisi 2

Model LR dibangun dengan mensubstitusikan nilai  $\beta_0, \beta_1, \dots, \beta_p$  pada

Persamaan 2.3, sehingga diperoleh

$$\pi_1(x_1) = \left( \frac{e^{-20,55+0,38x_1D_1+2,003x_1D_2+0,122x_3D_2-1,066x_5D_4+0,275x_6D_1-0,169x_7D_1+0,586x_8D_1+18,665x_9D_1+0,304x_{11}+0,273x_{12}D_1-0,123x_{13}D_3}}{1 + e^{-20,55+0,38x_1D_1+2,003x_1D_2+0,122x_3D_2-1,066x_5D_4+0,275x_6D_1-0,169x_7D_1+0,586x_8D_1+18,665x_9D_1+0,304x_{11}+0,273x_{12}D_1-0,123x_{13}D_3}} \right)$$

$$\pi_2(x_1) = \left( \frac{e^{-2,5+0,062x_2-0,099x_3D_1-0,523x_4D_1+0,223x_5D_1-0,376x_5D_2-0,437x_5D_3+0,262x_6D_2+2,131x_{10}D_1+0,432x_{13}D_1-0,175x_{13}D_2+0,599x_{13}D_4}}{1 + e^{-2,5+0,062x_2-0,099x_3D_1-0,523x_4D_1+0,223x_5D_1-0,376x_5D_2-0,437x_5D_3+0,262x_6D_2+2,131x_{10}D_1+0,432x_{13}D_1-0,175x_{13}D_2+0,599x_{13}D_4}} \right)$$

dengan  $\pi_1(x_1)$  adalah model LR yang terbentuk pada partisi pertama,

sedangkan  $\pi_2(x_1)$  adalah model LR yang terbentuk pada partisi kedua.

Berdasarkan model LR yang diperoleh dapat dilakukan transformasi logit untuk mempermudah interpretasi.

$$\begin{aligned} \text{Logit}[\pi_1(x_1)] &= \ln\left(\frac{\pi_1(x_1)}{1 - \pi_1(x_1)}\right) \\ &= -20,55 + 0,38x_1D_1 + 2,003x_1D_2 + 0,122x_3D_2 - 1,066x_5D_4 + 0,275x_6D_1 \\ &\quad - 0,169x_7D_1 + 0,586x_8D_1 + 18,665x_9D_1 + 0,304x_{11} + 0,273x_{12}D_1 \\ &\quad - 0,123x_{13}D_3 \end{aligned} \quad (4.1)$$

Berdasarkan model logit 4.1, dapat diinterpretasikan bahwa pengaruh variabel-variabel prediktor dapat diuraikan sebagai berikut:

- a. Variabel  $x_1D_1$ , yaitu berat badan bayi lahir lebih, memberikan pengaruh lebih tinggi sebesar 0,381 kali terhadap kematian bayi;
- b. Variabel  $x_1D_2$ , yaitu berat badan bayi lahir kurang, memberikan pengaruh lebih tinggi sebesar 2,003 kali terhadap kematian bayi;
- c. Variabel  $x_3D_2$ , yaitu pendidikan ibu tamatan  $\leq SD$ , memberikan pengaruh lebih tinggi sebesar 0,112 kali terhadap kematian bayi;
- d. Variabel  $x_5D_4$ , yaitu interval kelahiran  $< 18$  bulan, memberikan pengaruh lebih rendah sebesar 1,066 kali terhadap kematian bayi;
- e. Variabel  $x_6D_1$ , yaitu frekuensi melakukan pemeriksaan kehamilan tidak sesuai standar, memberikan pengaruh lebih tinggi sebesar 0,275 kali terhadap kematian bayi;
- f. Variabel  $x_7D_1$ , yaitu tempat persalinan tanpa fasilitas Kesehatan, memberikan pengaruh lebih rendah sebesar 0,169 kali terhadap kematian bayi;
- g. Variabel  $x_8D_1$ , yaitu jenis kelamin laki-laki, memberikan pengaruh lebih tinggi sebesar 0,586 kali terhadap kematian bayi;

- h. Variabel  $x_9D_1$ , yaitu tidak memberikan ASI, memberikan pengaruh lebih tinggi sebesar 18,665 kali terhadap kematian bayi;
- i. Variabel  $x_{11}$ , yaitu urutan kelahiran, memberikan pengaruh lebih tinggi sebesar 0,304 kali terhadap kematian bayi;
- j. Variabel  $x_{12}D_1$ , yaitu tempat tinggal perdesaan, memberikan pengaruh lebih tinggi sebesar 0,273 kali terhadap kematian bayi;
- k. Variabel  $x_{13}D_3$ , yaitu indeks kekayaan miskin, memberikan pengaruh lebih rendah sebesar 0,123 kali terhadap kematian bayi.

Selanjutnya, pada partisi kedua menunjukkan hasil transformasi logit sebagai berikut

$$\begin{aligned}
 \text{Logit}[\pi_2(x)] &= \ln\left(\frac{\pi_2(x)}{1 - \pi_2(x)}\right) \\
 &= -2,5 + 0,062x_2 - 0,099x_3D_1 - 0,523x_4D_1 + 0,223x_5D_1 - 0,376x_5D_2 \\
 &\quad - 0,437x_5D_3 + 0,262x_6D_2 + 2,131x_{10}D_1 + 0,432x_{13}D_1 - 0,175x_{13}D_2 \\
 &\quad + 0,599x_{13}D_4
 \end{aligned} \tag{4.2}$$

Berdasarkan model logit 4.2 dapat diinterpretasikan bahwa pengaruh variabel-variabel prediktor pada partisi kedua dapat diuraikan sebagai berikut:

- a. Variabel  $x_2$ , yaitu usia ibu saat melahirkan, memberikan pengaruh lebih tinggi sebesar 0,062 kali terhadap kematian bayi;
- b. Variabel  $x_3D_1$ , yaitu Pendidikan ibu lulusan SMP, memberikan pengaruh lebih rendah sebesar 0,099 kali terhadap kematian bayi;
- c. Variabel  $x_4D_1$ , yaitu pekerjaan ibu, memberikan pengaruh lebih rendah sebesar 0,523 kali terhadap kematian bayi;
- d. Variabel  $x_5D_1$ , yaitu interval kelahiran 36-47 bulan, memberikan pengaruh lebih tinggi sebesar 0,223 kali terhadap kematian bayi;

- e. Variabel  $x_5D_2$ , yaitu interval kelahiran 24-35 bulan, memberikan pengaruh lebih rendah sebesar 0,376 kali terhadap kematian bayi;
- f. Variabel  $x_5D_3$ , yaitu interval kelahiran 18-23 bulan, memberikan pengaruh lebih rendah sebesar 0,437 kali terhadap kematian bayi;
- g. Variabel  $x_6D_2$ , yaitu frekuensi tidak melakukan pemeriksaan kehamilan, memberikan pengaruh lebih tinggi sebesar 0,262 kali terhadap kematian bayi;
- h. Variabel  $x_{10}D_1$ , jenis kelahiran kembar, memberikan pengaruh lebih tinggi sebesar 2,131 kali terhadap kematian bayi;
- i. Variabel  $x_{13}D_1$ , yaitu indeks kekayaan keluarga kaya, memberikan pengaruh lebih tinggi sebesar 0,432 kali terhadap kematian bayi;
- j. Variabel  $x_{13}D_2$ , yaitu indeks kekayaan keluarga menengah, memberikan pengaruh lebih rendah sebesar 0,175 kali terhadap kematian bayi;
- k. Variabel  $x_{13}D_4$ , yaitu indeks kekayaan sangat miskin, memberikan pengaruh lebih tinggi sebesar 0,599 kali terhadap kematian bayi.

Seperti halnya pada *ensemble* pertama, pada *ensemble* kedua dan seterusnya dilakukan perlakuan yang serupa dalam memperoleh model LR. Model LR yang terbentuk pada masing-masing partisi dan *ensemble* digunakan untuk menghitung nilai probabilitas setiap pasangan data *testing* guna memprediksi kelas data *testing* itu sendiri. Berdasarkan Tabel 4.9 dan persamaan model logit 4.2 dapat dilihat variabel yang memberikan pengaruh paling tinggi pada kematian bayi adalah pemberian ASI.

#### 4.3.4 Nilai Prediksi

Nilai prediksi setiap prediksi diperoleh dengan mensubstitusikan data *testing* pada model yang sebelumnya sudah diperoleh. Perhitungan ini dilakukan pada masing-masing *ensemble*. Berikut merupakan hasil prediksi untuk masing-masing *ensemble*.

Tabel 4.10 Nilai Prediksi Setiap *Ensemble*

Data ke-	<i>Ensemble</i>						
	1	2	3	4	5	...	20
<b>1</b>	0.928	0.955	0.910	0.939	0.883	...	0.904
<b>11</b>	0.354	0.252	0.315	0.399	0.311	...	0.328
<b>16</b>	0.584	0.599	0.531	0.543	0.477	...	0.502
<b>21</b>	0.257	0.328	0.300	0.258	0.389	...	0.336
<b>33</b>	0.369	0.331	0.373	0.315	0.463	...	0.463
<b>54</b>	0.388	0.507	0.495	0.564	0.368	...	0.385
<b>55</b>	0.363	0.356	0.391	0.331	0.479	...	0.458
<b>78</b>	0.358	0.222	0.315	0.295	0.371	...	0.397
<b>84</b>	0.400	0.263	0.327	0.343	0.325	...	0.351
<b>89</b>	0.200	0.202	0.186	0.275	0.272	...	0.195
<b>91</b>	0.626	0.775	0.764	0.749	0.611	...	0.569
<b>103</b>	0.368	0.218	0.296	0.200	0.295	...	0.316
<b>106</b>	0.410	0.378	0.354	0.409	0.333	...	0.336
<b>123</b>	0.201	0.202	0.258	0.155	0.170	...	0.177
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<b>14955</b>	0.331	0.227	0.314	0.276	0.384	...	0.395

Tabel 4.10 Menunjukkan hasil prediksi pada data *testing* terhadap masing-masing *ensemble*. Pada data *testing* pertama, yaitu data pertama dan *ensemble* pertama menunjukkan nilai prediksi sebesar 0,928, sedangkan pada *ensemble* kedua menunjukkan nilai prediksi sebesar 0,955, hingga pada *ensemble* ke-20 menunjukkan nilai prediksi 0,904. Selanjutnya pada data *testing* kedua, yaitu data ke-11 dan *ensemble* pertama menunjukkan nilai prediksi sebesar 0,354, sedangkan pada *ensemble* kedua menunjukkan nilai prediksi sebesar 0,252, hingga pada *ensemble* ke-20 menunjukkan nilai prediksi sebesar 0,328.

### 4.3.5 *Threshold Optimal*

Kelebihan dari metode LORENS adalah pada nilai *threshold* yang digunakan. Umumnya pada LR digunakan nilai *threshold* 0,5, namun hal tersebut dirasa kurang baik apabila kelas 0 dan 1 tidak seimbang. *Threshold* optimal dapat diperoleh menggunakan persamaan 2.15 dengan perhitungan sebagai berikut.

$$Threshold = \frac{(\bar{y} + 0,5)}{2} = \frac{\left(\frac{4952}{13500} + 0,5\right)}{2} = \frac{0,36681 + 0,5}{2} = 0,4334073$$

Nilai *threshold* optimal yang diperoleh pada penelitian ini adalah 0,4334073. Nilai ini digunakan untuk menentukan kelas klasifikasi pada masing-masing nilai prediksi.

### 4.3.6 *Kelas Prediksi Klasifikasi*

Nilai prediksi pada penelitian ini dilakukan dengan mensubstitusikan data testing dalam LR. Nilai prediksi yang diperoleh dari rata-rata probabilitas setiap partisi pada masing-masing *ensemble*. Berdasarkan alur tersebut diperoleh hasil prediksi tiap *ensemble* sebagai berikut. Hasil prediksi akhir dilakukan dengan cara voting, yaitu dengan memilih keputusan prediksi terbanyak dari semua *ensemble*. Jika nilai probabilitas akhir lebih besar dari pada *threshold* optimal, maka data *testing* masuk kedalam kelas 1 yang merupakan bayi mati. Seperti itu pula sebaliknya, jika nilai probabilitas akhir kurang dari nilai *threshold* optimal, maka data *testing* masuk ke dalam kelas 0 yang merupakan bayi hidup. Berikut merupakan nilai rata-rata probabilitas dan keputusan kelas klasifikasi pada model yang terbentuk dari 3 partisi dengan *threshold* optimal yaitu 0,4334073.

Tabel 4.11 Nilai Rata-Rata Probabilitas dan Keputusan Klasifikasi

Data ke-	Ensemble							Rata-rata	Perbandingan <i>Threshold</i>	Keputusan
	1	2	3	4	5	...	20			
<b>1</b>	0.928	0.955	0.910	0.939	0.883	...	0.904	0.910	>0,4334	1
<b>11</b>	0.354	0.252	0.315	0.399	0.311	...	0.328	0.329	<0,4334	0
<b>16</b>	0.584	0.599	0.531	0.543	0.477	...	0.502	0.553	>0,4334	1
<b>21</b>	0.257	0.328	0.300	0.258	0.389	...	0.336	0.291	<0,4334	0
<b>33</b>	0.369	0.331	0.373	0.315	0.463	...	0.463	0.352	<0,4334	0
<b>54</b>	0.388	0.507	0.495	0.564	0.368	...	0.385	0.457	>0,4334	1
<b>55</b>	0.363	0.356	0.391	0.331	0.479	...	0.458	0.375	<0,4334	0
<b>78</b>	0.358	0.222	0.315	0.295	0.371	...	0.397	0.312	<0,4334	0
<b>84</b>	0.400	0.263	0.327	0.343	0.325	...	0.351	0.323	<0,4334	0
<b>89</b>	0.200	0.202	0.186	0.275	0.272	...	0.195	0.216	<0,4334	0
<b>91</b>	0.626	0.775	0.764	0.749	0.611	...	0.569	0.693	>0,4334	1
<b>103</b>	0.368	0.218	0.296	0.200	0.295	...	0.316	0.270	<0,4334	0
<b>106</b>	0.410	0.378	0.354	0.409	0.333	...	0.336	0.369	<0,4334	0
<b>123</b>	0.201	0.202	0.258	0.155	0.170	...	0.177	0.185	<0,4334	0
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
<b>14955</b>	0.331	0.227	0.314	0.276	0.384	...	0.395	0.319	<0,4334	0

Tabel 4.11 menunjukkan pada rata-rata nilai probabilitas pertama pada data *testing* menghasilkan nilai 0,910. Dibandingkan dengan *threshold* optimal maka nilai tersebut lebih besar, artinya keputusan klasifikasinya adalah masuk ke dalam kelas 1. Pada data *testing* kedua nilai rata-rata probabilitas adalah 0,329, artinya nilai rata-rata probabilitas kedua masuk ke dalam kelas 0. Pada data *testing* terakhir yaitu nilai rata-rata probabilitas lebih dari nilai *threshold* yaitu sebesar 0,319, sehingga semua keputusan klasifikasi masuk ke dalam kelas 0.

#### 4.3.7 Menghitung Ketepatan Hasil Klasifikasi

Perhitungan ketepatan hasil klasifikasi dapat diperoleh melalui hasil perhitungan dari tabulasi silang yang telah terdiri dari *True Positive* (TP), *False Positive* (FP), *True Negative* (TN,) dan *False Negative* (FN). Berikut merupakan tabel tabulasi silang dengan dua partisi dan 20 *ensemble*.

Tabel 4.12 Hasil Tabulasi Silang Klasifikasi

	Kelas <i>actual</i>		
		$p(+)$	$n(-)$
Kelas Prediksi	$p(+)$	393	157
	$n(-)$	121	829

Tabel 4.12 menunjukkan bahwa nilai *True Positive* adalah 393, artinya 393 kasus mati pada bayi yang tepat diprediksi mati. Nilai *False Negative* adalah 157, artinya 157 kasus mati pada bayi yang terprediksi hidup. Nilai *True Negative* adalah 829, artinya terdapat 829 kasus bayi hidup yang tepat terprediksi hidup. Sedangkan nilai *False Positive*-nya adalah 121, artinya terdapat 121 kasus bayi hidup yang terprediksi mati. Berdasarkan persamaan 2.16 s.d. 2.18 diperoleh nilai *sensitivity*, *specificity*, dan *accuracy* dari klasifikasi LORENS pada AKB adalah sebagai berikut.

$$\begin{aligned} \text{Sensitivity} &= \frac{393}{393 + 121} \times 100\% \\ &= 76,45\% \end{aligned}$$

$$\begin{aligned} \text{Specificity} &= \frac{829}{157 + 829} \times 100\% \\ &= 84,08\% \end{aligned}$$

$$\begin{aligned} \text{Accuracy} &= \frac{393 + 829}{393 + 157 + 121 + 829} \times 100\% \\ &= 81,47\% \end{aligned}$$

Hasil yang ditunjukkan pada perhitungan *sensitivity* menunjukkan hasil yaitu 76,45%, *specificity* adalah 84,08%, dan *accuracy* adalah 81,47%. Artinya tingkat *sensitivity*, *specificity*, maupun *accuracy* model LORENS terhadap AKB memiliki nilai yang baik. Selain untuk menentukan nilai *sensitivity*, *specificity*, dan *accuracy*, tabel tabulasi dapat digunakan untuk

mencari nilai kesalahan dalam klasifikasi. Kesalahan klasifikasi dapat diperoleh menggunakan persamaan 2.19 s.d. 2.21 yang terdiri dari *False Positive Rate* (FPR), *False Negative Rate* (FNR), dan *error*.

$$\begin{aligned} FPR &= \frac{157}{393 + 121} \times 100\% \\ &= 30,54\% \end{aligned}$$

$$\begin{aligned} FNR &= \frac{121}{393 + 121} \times 100\% \\ &= 22\% \end{aligned}$$

$$\begin{aligned} error &= \frac{157 + 121}{393 + 157 + 121 + 829} \times 100\% \\ &= 18,53\% \end{aligned}$$

Hasil kesalahan klasifikasi yang diperoleh menunjukkan hasil *False Positive Rate* (FPR) adalah 30,54% dan *False Negative Rate* (FNR) adalah 22%. Sedangkan untuk tingkat *error* klasifikasi menunjukkan 18,53%, artinya dalam proses analisis klasifikasi LORENS memiliki *error* klasifikasi sebesar 18,53%.

#### 4.4 Evaluasi Performa Model

Metode evaluasi kebaikan model klasifikasi yang digunakan dalam penelitian ini adalah *cross validation*. Semua data berperan sebagai data *training* sekaligus data *testing* sesuai dengan gilirannya masing-masing. Penelitian ini membagi data dalam 10 *folds* di mana masing-masing *fold* berperan sebagai data *training* dan data *testing*. Ketika *fold* pertama digunakan sebagai data *testing* maka *fold* kedua hingga kesepuluh digunakan sebagai data *training*. Ketika *fold* kedua digunakan sebagai data *testing*, maka *fold* pertama dan *fold* ketiga hingga

ke-10 digunakan sebagai data *training*. Pola yang sama dilakukan hingga semua *fold* sudah berperan sebagai data *training* sekaligus data *testing*.

Partisi yang digunakan dalam langkah ini ditentukan dengan melihat nilai akurasi terbaik dari setiap partisinya. Berikut merupakan perbandingan nilai akurasi pada setiap partisi.

Tabel 4.13 Partisi Optimal *Cross Validation*

<b>Banyak Partisi</b>	<b>Akurasi (%)</b>
2	79,64
3	76,31
4	74,97
5	74,38
6	73,23
⋮	⋮
22	63,51

Tabel 4.13 menunjukkan bahwa nilai akurasi terbaik adalah ketika data dibagi ke dalam 2 partisi yaitu dengan nilai akurasi 79,64%. Sedangkan pada partisi ketiga hingga ke-22 selalu mengalami penurunan akurasi.

Tabel 4.14 Hasil *Threshold Optimal Cross Validation*

<b><i>Fold</i></b>	<b><i>Threshold Optimal</i></b>	<b><i>Fold</i></b>	<b><i>Threshold Optimal</i></b>
1	0,43289	6	0,43233
2	0,43304	7	0,4330
3	0,43444	8	0,434
4	0,43337	9	0,43474
5	0,43311	10	0,43304

Tabel 4.14 menunjukkan hasil masing-masing *threshold* optimal pada masing-masing *fold* yang telah dibagi. Selanjutnya untuk mengetahui tingkat ketepatan evaluasi performa model klasifikasi dapat diperoleh menggunakan tabel tabulasi silang. Berikut merupakan hasil tabulasi silang klasifikasi pada data AKB untuk 2 partisi.

Tabel 4.15 Tabulasi Silang Hasil Klasifikasi *Cross Validation*

	Kelas <i>actual</i>		
		$p(+)$	$n(-)$
Kelas	$p(+)$	3783	1719
Prediksi	$n(-)$	1335	8163

Berdasarkan Tabel 4.15 diketahui bahwa nilai *True Positive* adalah 3783, artinya 3783 kasus mati pada bayi yang tepat diprediksi mati. Nilai *False Negative* adalah 1719, artinya 1719 kasus mati pada bayi yang terprediksi hidup. Nilai *True Negative* adalah 8163, artinya terdapat 8163 kasus bayi hidup yang tepat terprediksi hidup. Sedangkan nilai *False Positive*-nya adalah 1335, artinya terdapat 1335 kasus bayi hidup yang terprediksi mati. Sehingga berdasarkan tabel tabulasi diatas diperoleh nilai *sensitivity*, *specificity*, dan *accuracy* untuk evaluasi model berikut.

$$\begin{aligned} \text{Sensitivity} &= \frac{3783}{3783 + 1335} \times 100\% \\ &= 73,92\% \end{aligned}$$

$$\begin{aligned} \text{Specificity} &= \frac{8163}{1719 + 8163} \times 100\% \\ &= 82,6\% \end{aligned}$$

$$\begin{aligned} \text{Accuracy} &= \frac{3783 + 8163}{3783 + 1719 + 1335 + 8163} \times 100\% \\ &= 79,64\% \end{aligned}$$

Hasil yang ditunjukkan pada perhitungan *sensitivity* menunjukkan hasil yaitu 73,92%, *specificity* adalah 82,6%, dan *accuracy* adalah 79,64%. Artinya tingkat *sensitivity*, *specificity*, maupun *accuracy* model LORENS terhadap AKB memiliki nilai yang cukup baik

#### 4.5 Kajian Keislaman dengan Hasil Penelitian

Berdasarkan hasil penelitian angka kematian bayi menggunakan *logistic regression ensemble* (LORENS) diketahui bahwa nilai koefisien dari pemberian ASI kepada bayi memberikan pengaruh positif terhadap kematian bayi dengan nilai yang cukup tinggi di setiap model logitnya. Hal ini menunjukkan betapa pemberian ASI merupakan salah satu faktor yang penting dalam menjaga Kesehatan bayi. Pemberian ASI dalam Islam tercantum dalam QS. Al-Baqarah ayat 233 sebagai berikut:

*“Para Ibu hendaklah menyusukan anak-anaknya selama dua tahun penuh, yaitu bagi yang ingin menyempurnakan penyusuan, dan kewajiban ayah memberi makan dan pakaian kepada para ibu dengan cara ma’ruf. Seseorang tidak dibebani melainkan menurut kadar kesanggupannya. Janganlah seorang ibu menderita kesengsaraan karena anaknya dan seorang ayah karena anaknya., dan warispun berkewajiban demikian. Apabila keduanya ingin menyapih (sebelum dua tahun) dengan kerelaan keduanya dan permusyawaratan, maka tiada dosa atas keduanya, dan jika kamu ingin anakmu disusukan oleh orang lain, maka tiada dosa bagimu apabila kamu memberikan pembayaran menurut yang patut. Bertakwalah kamu kepada Allah dan ketahuilah Allah maha melihat apa yang kamu kerjakan”* (Kementerian Agama RI, 2019)

Menurut Tafsir Tahlili Kementerian Agama RI (2019), surah Al Baqarah ayat 233 memberikan penjelasan bahwa Setiap ibu (meskipun ia janda) berkewajiban menyusui anaknya sampai anak itu mencapai usia dua tahun. Tidak mengapa kalau masa susuan itu kurang dari masa tersebut apabila kedua ibu-bapak memandang ada maslahatnya. Demikian pula setiap bapak berkewajiban untuk memenuhi kebutuhan para ibu baik sandang maupun pangan sesuai dengan kebutuhannya. Ibu laksana wadah bagi anak sedang bapak sebagai pemilik wadah itu. Maka sudah sewajarnya bapak berkewajiban memberi nafkah kepada orang yang di bawah tanggung jawabnya dan memelihara serta merawat miliknya.

Allah mewajibkan kepada ibu menyusui bayinya, karena Air Susu Ibu (ASI) mempunyai pengaruh yang bears kepada anaknya. Dari hasil penelitian para ahli medis menunjukkan bahwa ASI terdiri dari saripati yang benar-benar murni. ASI juga merupakan makanan yang paling baik untuk bayi, dan tidak ragukan lagi oleh para ahli gizi. Di samping itu dengan fitrah kejadiannya memiliki rasa kasih sayang yang mendalam sehingga penyusuan langsung dari ibu, berhubungan erat dengan perkembangan jiwa dan mental anak. Oleh karenanya kurang tepat tindakan sementara para ibu yang tidak mau menyusui anaknya hanya karena kepentingan pribadinya, umpamanya, untuk memelihara kecantikan. Padahal ini bertentangan dengan fitrahnya sendiri dan secara tidak langsung ia kehilangan kesempatan untuk membina dasar hubungan keibuan dengan anaknya sendiri dalam bidang emosi.

Demikianlah pembagian kewajiban kedua orang tua terhadap bayinya yang diatur oleh Allah swt. Sementara itu diberi pula keringanan terhadap kewajiban, umpama kesehatan ibu terganggu atau seorang dokter mengatakan tidak baik bila disusukan oleh ibu karena suatu hal, maka tidak mengapa kalau anak mendapat susuan atau makanan dari orang lain.

Demikian juga apabila bapak tidak mempunyai kesanggupan melaksanakan kewajibannya karena miskin maka ia boleh melaksanakan kewajibannya sesuai dengan kesanggupannya. Keringanan itu membuktikan bahwa anak tidak boleh dijadikan sebab adanya kemudharatan, baik terhadap bapak maupun terhadap ibu. Oleh karenanya, kewajiban tersebut tidak mesti berlaku secara mutlak sehingga mengakibatkan kemudharatan bagi keduanya. Salah satu pihak tidak boleh memudharatkan pihak lain dengan menjadikan anak

sebagai kambing hitamnya. Umpamanya karena ibu mengetahui bahwa bapak berkewajiban memberi nafkah maka ia melakukan pemerasan dengan tidak menyusui atau merawat si bayi tanpa sejumlah biaya tertentu. Atau bapak sangat kikir dalam memberikan nafkah sehingga ibu menderita karenanya.

Berdasarkan surah Al-Baqarah ayat 233 ini menunjukkan bahwa sangatlah penting pemberian ASI pada bayi oleh Ibu. Bahkan Allah SWT menganjurkannya dalam surah ini. Pada model yang dihasilkan pada *ensemble* pertama hingga *ensemble* ke-20 menunjukkan hasil model yang menerangkan bahwa pemberian ASI memberikan pengaruh jauh lebih besar diantara variabel-variabel lainnya. Hal ini menunjukkan bahwa betapa pentingnya memberikan kepada bayi sebelum berusia dua tahun.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan dari hasil analisis dan pembahasan klasifikasi angka kematian bayi (AKB) diperoleh kesimpulan sebagai berikut.

1. Analisis LORENS yang dilakukan pada angka kematian bayi (AKB) memberikan hasil klasifikasi terbaik dengan pembagian data *training* dan data *testing* adalah 90%:10%. Partisi yang optimal yang terbentuk adalah sebanyak 2 partisi, dan banyak *ensemble* adalah 20 *ensemble*. Selain itu, *threshold* optimal yang diperoleh dalam penelitian ini adalah 0,4334073. Model LR yang diperoleh terbentuk berdasarkan banyaknya partisi dan *ensemble* yang digunakan. Berdasarkan partisi dan *ensemble* tersebut terbentuklah model LR yaitu 2 model dari setiap partisi, artinya untuk 20 *ensemble* terbentuk 40 model LR. Berdasarkan model LR yang dibangun diketahui bahwa variabel pemberian ASI merupakan variabel yang memiliki pengaruh terbesar terhadap kematian pada bayi.
2. Tingkat *accuracy* pada klasifikasi angka kematian bayi (AKB) menggunakan LORENS didapatkan hasil sebesar 81,47%. Artinya model yang terbentuk mampu melakukan klasifikasi dengan tingkat ketepatan sebesar 81,47%. Sama halnya dengan nilai *accuracy*, nilai *sensitivity* yang diperoleh adalah 76,45%. Selain itu, diperoleh nilai *specificity* adalah 84,08%. Klasifikasi LORENS dengan evaluasi performa model berupa *cross validation* digunakan pembagian data dalam 10 *folds*. Setiap *fold*

digunakan sebagai data *training* sekaligus data *testing*. Penggunaan 10 *folds cross validation* memperoleh nilai *accuracy* adalah 79,64%, nilai *specificity* 82,6%, dan nilai *sensitivity* adalah 73,92%.

## 5.2 Saran

Saran oleh penulis terhadap penelitian selanjutnya adalah penggunaan kombinasi baru pada perbandingan data *training* dan *testing* yang memiliki nilai *accuracy* terbaik. Selain itu, Pada kombinasi *ensemble* dan partisi bisa dicarikan *script* untuk mempermudah penemuan kombinasi terbaik, sehingga tidak perlu mencari satu-persatu pada setiap perbandingan.

## DAFTAR PUSTAKA

- Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). John Wiley & Sons.
- Ahn, H., Moon, H., Fazzari, M. J., Lim, N., Chen, J. J., & Kodell, R. L. (2007). *Classification by ensembles from random partitions of high-dimensional data*. 51, 6166–6179. <https://doi.org/10.1016/j.csda.2006.12.043>
- Al-Bukhari, A. A. M. bin I. (2011). *Ensiklopedia hadits: Shahih Al-Bukhari 1*. Almahira.
- Al-Qazwiri, I. M. A. A. M. bin Y. (2021). Hadis Sunan Ibnu Majah Jilid 2. In Tim Turas Nabawi (Ed.), *Maktabah Syamilah* (2nd ed.). Turats Nabawi Press.
- Alvaro, R., Christianingrum, R., & Riyono, T. (2019). Analisis RKP dan Pembicaraan Pendahuluan APBN. *Angewandte Chemie International Edition*, 6(11), 951–952., 3(2), 1.
- Asfihani, A. (2015). *Prediksi Pembelotan Konsumen Software Antivirus Binary Logistic Regression dan Logistic Regression Ensembles*.
- Azizah, L. N. (2013). *Pengujian Signifikansi Model Geographically Weighted Regression (GWR) dengan Statistik Uji F dan Uji T*. UIN Maulana Malik Ibrahim Malang.
- Baudi, S. (2019). *Analisis Tingkat Kerugian Material Akibat Kecelakaan Lalu Lintas Dengan Menggunakan Dummy Variable di Provinsi Riau Tahun 2013-2017*. Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Catal, C. (2012). *Performance Evaluation Metrics for Software Fault Prediction Studies*. 9(4), 193–206.
- Choirunnisa, S. (2019). *Metode Hibrida Oversampling dan Undersampling untuk Menangani Ketidakseimbangan Data Kegagalan Akademik Universitas XYZ*.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression second Edition*. Wiley Series in Probability and Statistics.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression Third Edition*. In *Technometrics* (Vol. 38, Issue 2). <https://doi.org/10.2307/1270433>
- Kementerian Agama RI. (2019). *Quran Kemenag*. <https://quran.kemenag.go.id/>
- Lim, N. (2007). *Classification by Ensembles from Random Partitions using Logistic Regression Models*. Stony Brook University.
- Lim, N., Ahn, H., Moon, H., & Chen, J. J. (2010). *Classification of High-Dimensional Data With Ensemble of Logistic Regression Model*. 160–171. <https://doi.org/10.1080/10543400903280639>
- Melasasi, J. N. (2015). *Klasifikasi Enzim Pada Database Dud-E Dengan Metode Logistic Regression Ensemble (LORENS) dengan Metode Logistic Regression*. Institut Teknologi Sepuluh Nopember.

- Ratnawati, S. (2020). *Penggunaan Metode Logistic Regression Ensemble (LORENS) Pada Klasifikasi Leukimia Akut*. Universitas Islam Bandung.
- Ratnawati, S., & Sunendiari, S. (2021). *Penggunaan Metode Logistic Regression Ensemble ( LORENS ) pada Klasifikasi Leukemia Akut*. 7, 56–63. <https://doi.org/10.29313/.v7i1.25555>
- Widhianingsih, T. D. A. (2018). *Klasifikasi Data Berdimensi Tinggi dengan Metode Ensemble berbasis Regresi Logistik dalam permasalahan Drug Discovery*. Institut Teknologi Sepuluh Nopember.
- Widhianingsih, T. D. A., Kuswanto, H., & Prastyo, D. D. (2020). Logistic Regression Ensemble (LORENS) Applied to Drug Discovery. *Matematika*, 36(1), 43–49. <https://doi.org/10.11113/matematika.v36.n1.1197>
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques 3rd Edition*.

## LAMPIRAN

Lampiran 1. Data Angka Kematian Bayi 2017

No	y	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	x <sub>8</sub>	x <sub>9</sub>	x <sub>10</sub>	x <sub>11</sub>	x <sub>12</sub>	x <sub>13</sub>
1	0	2	38	2	1	1	0	0	1	1	1	6	1	4
2	0	1	28	2	0	2	0	0	1	1	0	4	1	2
3	0	1	39	2	0	3	0	0	1	1	0	5	1	4
4	0	0	35	2	0	2	0	1	1	0	0	8	1	3
5	0	1	30	2	1	1	0	0	1	0	0	3	1	4
6	0	0	28	0	0	1	0	0	0	0	0	2	1	3
7	0	0	38	0	0	1	0	0	0	0	0	2	1	3
8	0	1	26	2	0	1	0	0	1	0	0	3	1	4
9	0	1	38	2	1	0	0	0	1	0	0	3	1	3
10	0	0	26	1	1	2	0	1	1	1	0	2	1	3
11	0	1	24	1	0	2	0	1	1	1	0	2	1	1
12	0	1	39	1	1	4	1	0	1	1	0	3	1	1
13	0	1	29	2	1	1	0	0	1	0	0	2	1	4
14	0	2	30	1	1	3	0	0	1	1	0	4	1	4
15	0	1	36	2	1	0	0	0	0	0	0	3	1	4
16	0	1	42	2	1	2	0	0	1	1	0	8	1	3
17	0	1	36	2	1	4	0	1	0	1	0	2	1	4
18	0	1	42	2	1	4	0	0	1	1	0	8	1	4
19	0	1	30	1	0	2	0	0	0	1	0	3	0	2
20	0	1	27	1	0	3	0	0	0	1	0	2	0	3
21	0	1	28	1	1	4	0	0	1	1	0	2	0	4
22	0	1	35	2	1	0	0	0	0	0	0	4	0	2
23	0	1	31	0	0	0	0	0	0	0	0	2	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<b>9626</b>	0	1	21	1	1	4	1	0	1	1	0	2	0	1

Lampiran 2. Data Hasil *Dummy*

$x_1D_1$	$x_1D_2$	$x_2$	$x_3D_1$	$x_3D_2$	$x_4D_1$	$x_5D_1$	$x_5D_2$	$x_5D_3$	$x_5D_4$	$x_6D_1$	$x_6D_2$
0	1	38	0	1	1	1	0	0	0	0	0
1	0	28	0	1	0	0	1	0	0	0	0
1	0	39	0	1	0	0	0	1	0	0	0
0	0	35	0	1	0	0	1	0	0	0	0
1	0	30	0	1	1	1	0	0	0	0	0
0	0	28	0	0	0	1	0	0	0	0	0
0	0	38	0	0	0	1	0	0	0	0	0
1	0	26	0	1	0	1	0	0	0	0	0
1	0	38	0	1	1	0	0	0	0	0	0
0	0	26	1	0	1	0	1	0	0	0	0
1	0	24	1	0	0	0	1	0	0	0	0
1	0	39	1	0	1	0	0	0	1	1	0
1	0	29	0	1	1	1	0	0	0	0	0
0	1	30	1	0	1	0	0	1	0	0	0
1	0	36	0	1	1	0	0	0	0	0	0
1	0	42	0	1	1	0	1	0	0	0	0
1	0	36	0	1	1	0	0	0	1	0	0
1	0	42	0	1	1	0	0	0	1	0	0
1	0	30	1	0	0	0	1	0	0	0	0
1	0	30	0	1	1	1	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	0	31	0	1	1	1	0	0	0	0	0

Lampiran 3. Lanjutan Hasil Data *Dummy*

$x_7D_1$	$x_8D_1$	$x_9D_1$	$x_{10}D_1$	$x_{11}$	$x_{12}D_1$	$x_{13}D_1$	$x_{13}D_2$	$x_{13}D_3$	$x_{13}D_4$
0	1	1	1	6	1	0	0	0	1
0	1	1	0	4	1	0	1	0	0
0	1	1	0	5	1	0	0	0	1
1	1	0	0	8	1	0	0	1	0
0	1	0	0	3	1	0	0	0	1
0	0	0	0	2	1	0	0	1	0
0	0	0	0	2	1	0	0	1	0
0	1	0	0	3	1	0	0	0	1
0	1	0	0	3	1	0	0	1	0
1	1	1	0	2	1	0	0	1	0
1	1	1	0	2	1	1	0	0	0
0	1	1	0	3	1	1	0	0	0
0	1	0	0	2	1	0	0	0	1
0	1	1	0	4	1	0	0	0	1
0	0	0	0	3	1	0	0	0	1
0	1	1	0	8	1	0	0	1	0
1	0	1	0	2	1	0	0	0	1
0	1	1	0	8	1	0	0	0	1
0	0	1	0	3	0	0	1	0	0
0	1	0	0	3	1	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	1	0	4	0	0	0	1	0

Lampiran 4. *Script* LORENS

```

lr.cerp <- function(y,x,nens,fixsize=NULL,fixthres=NULL,search=F){
  # initialization
  set.seed(256)
  options(warn=-1)
  if(sum(is.na(xtrain_dummy))>0) stop("missing value is found")
  if(sum(is.na(ytrain_dummy))>0) stop("missing value is found")
  y <- as.data.frame(ytrain_dummy)
  x <- as.data.frame(xtrain_dummy)
  num_pred <- ncol(x)
  num_obs <- nrow(x)
  pos_rate <- sum(y)/num_obs

  # parameter search or default option
  if(search==T) {
    #search.thre_size belum bisa di buka
    optimal <- search.thre_size(y,x,"lr")
    optsize <- optimal$size
    opthreshold <- optimal$threshold
  }else {
    #yg true (search!=T)
    if(is.null(fixsize)) fixsize <-2
    if(is.null(fixthres)) fixthres <- (pos_rate+.5)/2
    optsize <- fixsize
    opthreshold <- fixthres
  }
  #main body
  ptss <- floor(seq(1,optsize+.999,length.out=num_pred))
  fitted <- NULL;
  predicted <- NULL;
  cname <- NULL;
  coef.table<-matrix(0,num_pred,nens);
  partition.table <- matrix(0,num_pred,nens);
  intc <-matrix(0,optsize,nens);
  probability <- rep(0,num_obs)
  for (i in 1:nens){
    cname <- c(cname, paste("ens",i,sep=""))
    rand_pred <- sample(ptss)
    partition.table[,i] <- rand_pred
    avg_fit <- rep(0,num_obs)
    for(j in 1:optsize){
      smp_dt<- cbind(y,x[,rand_pred==j])
      intlr<- glm(kematian_bayi~.,data=smp_dt, family=binomial())
      coef.vector<-intlr$coefficient
      coef.vector[is.na(coef.vector)]<-0
      intc[j,i]<-coef.vector[1];
      coef.vector<-coef.vector[-1]
      coef.table[rand_pred==j,i]<- coef.vector
      avg_fit<-avg_fit+intlr$fitted.values
    }
    fitted<-cbind(fitted, avg_fit/optsize, deparse.level=0)
    probability<-probability+(avg_fit/optsize)/nens
  }
}

```

```

}
learning.decision <- ens.voting(fitted,opthreshold)$final.vote
colnames(fitted) <- cname
colnames(intc) <- cname
colnames(coef.table) <- cname;
rownames(coef.table) <- colnames(x)
colnames(partition.table) <- cname;
rownames(partition.table) <- colnames(x)
return(list(fitted=fitted,probability=probability,learning.decision=learning.decision,
           partition.table=partition.table,coef.table=coef.table,intercept=intc,
           number.ensemble=nens,optimal.size=optsize,optimal.threshold=opthreshold))
}

### lr.cerp.predict applies lr.cerp model to new data(test set) similar as predict.lm
function.
### lr.cerp.object is required and built from lr.cerp function.
### xtest is also required and should be same format as x in lr.cerp function.
### ytest is optional if you want to check the accuracy lr.cerp.predict <-
function(lr.cerp.object,xtest,ytest=NULL) { # initialization
lr.cerp.predict <- function(lr.cerp.object,xtest,ytest_dummy=NULL) {
# initialization
options(warn=-1)
if(sum(is.na(xtest_dummy))>0) stop("missing value is found")
if(sum(is.na(ytest_dummy))>0) stop("missing value is found")
xtest <- as.data.frame(xtest_dummy)
num_obs <- nrow(xtest)
nens <- lr.cerp.object$number.ensemble
optsize <- lr.cerp.object$optimal.size
opthreshold <- lr.cerp.object$optimal.threshold

# main body
cname <- NULL; test.decision <- NULL; fitted <- NULL; probability <- rep(0,num_obs)
xtest <- xtest[,rownames(lr.cerp.object$partition.table)]
for (i in 1:nens) {
  avg_fit <- rep(0,num_obs)
  cname <- c(cname, paste("ens",i,sep=""))
  curmod <- lr.cerp.object$partition.table[,i]
  for(j in 1:optsize) {
    intc <- lr.cerp.object$intercept[j,i]
    wrkmat <- xtest[,curmod==j]
    cvec <- lr.cerp.object$coef.table[curmod==j,i]
    int_vl <- as.matrix(wrkmat)%*%cvec
    int_vl <- int_vl + intc
    int_vl[int_vl>=709] <- 709
    avg_fit <- avg_fit + exp(int_vl)/(1+exp(int_vl))
  }
  fitted <- cbind(fitted,avg_fit/optsize,deparse.level=0)
  probability <- probability+(avg_fit/optsize)/nens
}
test.decision <- ens.voting(fitted, opthreshold, ytest_dummy)
colnames(fitted) <- cname

return(list(fitted=fitted,probability=t(probability),decision=test.decision$final.vote,

```

```

optimal.size=optsize,optimal.threshold=opthreshold,decision.table=test.decision$twobytw
wo))
}

lr.cerp.cv <- function(y,x,nens,v_fold,fixsize=NULL,fixthres=NULL,search=F, alpha=0,
type.measure="auc", nfold.reg.lr=5){
  set.seed(256)
  options(warn=-1)
  if(sum(is.na(yover))>0) stop("missing value is found")
  if(sum(is.na(xover))>0) stop("missing value is found")
  y <- as.data.frame(yover)
  x <- as.data.frame(xover)
  dim(y)
  num_obs <- nrow(y)
  rand_obs <- sample(1:num_obs)
  obs_rem <- num_obs%%v_fold
  obs_div <- (num_obs-obs_rem)/v_fold

  # main body
  probability <- rep(0,num_obs); predicted <- rep(0,num_obs); tbtable <- matrix(0,2,2)
  part_size.list<-NULL; threshold.list<-NULL

  for(i in 1:v_fold) {
    if(i<=obs_rem) {head1<-(i-1)*(obs_div+1)+1;tail1<-i*(obs_div+1);
    } else {head1<-(i-1)*obs_div+obs_rem+1;tail1<-i*obs_div+obs_rem;}
    test_seq<-rand_obs[head1:tail1]
    learn_seq<-rand_obs[-c(head1:tail1)]
    ylearn<-y[learn_seq,];xlearn<-x[learn_seq,];
    xtest<-x[test_seq,];ytest<-y[test_seq,];
    #lr.cerp belum bisa dibaca
    mid_rs<-lr.cerpcv(ylearn,xlearn,nens=10,fixsize=22,fixthres=NULL,search=F)
    pred_rs<-lr.cerp.predictcv(mid_rs,xtest,ytest)
    predicted[test_seq]<-pred_rs$decision
    for(j in 1:nens) probability[test_seq]<-probability[test_seq]+pred_rs$fitted[,j]/nens
    tbtable<-tbtable+pred_rs$decision.table
    part_size.list<-c(part_size.list,mid_rs$optimal.size)
    threshold.list<-c(threshold.list,mid_rs$optimal.threshold)
  }
  return((list(probability=probability,predicted=predicted,partition.size.list=part_size.list,
threshold.list=threshold.list,decision.table=tbtable)))
}

### internal functions
ens.voting <- function (tot_res,threshold,y=NULL){
  nens<-ncol(tot_res);
  nobs<-nrow(tot_res)
  if (!is.null(y)) {real_pos<-sum(y);real_neg<-nobs-real_pos}
  tot_res[tot_res>=threshold] <- 1; tot_res[tot_res<threshold] <- 0
  final.vote <- rep(0,nobs)
  for(i in 1:nobs)
    final.vote[i] <- mean(tot_res[i,])
  final.vote[final.vote>=0.5] <- 1; final.vote[final.vote<0.5] <- 0
}

```

```

twobytwo <- NULL
if (!is.null(y)) {
  real_pred_pos <- sum(final.vote==y&y==1)
  real_pred_neg <- sum(final.vote==y&y==0)
  real_pos_pred_neg <- real_pos - real_pred_pos
  real_neg_pred_pos <- real_neg - real_pred_neg
  twobytwo
  <-
rbind(c(real_pred_pos,real_pos_pred_neg),c(real_neg_pred_pos,real_pred_neg))
  rownames(twobytwo) <- c("real.pos","real.neg")
  colnames(twobytwo) <- c("pred.pos","pred.neg")
}
return(list(final.vote=final.vote,twobytwo=twobytwo))
}

search.thre_size <- function (y,x,method) {
  nprd <- ncol(x);nobs <- nrow(x);orate <- sum(y)/nobs
  szseq <- NULL; int_fits <- NULL
  initseed <- c(2,3,4,5,6,7,8,9,10,12)
  for (i in initseed) {
    ipt<-i*nprd/nobs
    ipt<-floor(ipts)
    if (ipts%%2==0) ipt<-ipts+1
    if (szseq[length(szseq)]!=ipts||is.null(szseq)) {
      szseq <- c(szseq,ipts)
      #cv.vit belum jalan
      int_fits <- cbind(int_fits,cv.fit(y,x,ipts,method))
    }
  }
  nsrsz <- length(szseq)
  add_fits<-NULL;addsz<-NULL
  if(orate>=.5) {iseq<-seq(.5,orate,.02)
} else {iseq<-seq(.5,orate,-.02);iseq<-rev(iseq)}
  nbis<-length(iseq)
  szfth<-rep(0,nbis);acfth<-rep(0,nbis)
  for(j in 1:nbis) {
    acseq<-rep(0,nsrsz)
    for(k in 1:nsrsz) {
      tmpf<-rep(0,nobs)
      tmpf[int_fits[,k]>=iseq[j]]<-1;tmpf[int_fits[,k]<iseq[j]]<-0
      acseq[k]<-sum(tmpf==y)/nobs
    }
  }
  nbst<-sum(acseq==max(acseq));scol<-seq(1:nsrsz)
  if(nbst==1) {nthc<-scol[acseq==max(acseq)]
} else {
  tmpcol<-scol[acseq==max(acseq)]
  nthc<-tmpcol[round(nbst/2)]
}
  #true
  if(nthc==1) {
    upts<-szseq[nthc+1];lpts<-szseq[nthc]
    utfac<-acseq[nthc+1];ltfac<-acseq[nthc]
    while(lpts!=upts) {
      mpts<-(lpts+upts)/2

```

```

mpts<-floor(mpts)
if(mpts%%2==0) mpts<-mpts+1
if(mpts==upts) break
if(length(addsz)==0) {
  mtf<-cv.fit(y,x,mpts,method)
  addsz<-c(addsz,mpts);add_fits<-cbind(add_fits,mtf)
}
else if(sum(addsz==mpts)==0) {
  mtf<-cv.fit(y,x,mpts,method)
  addsz<-c(addsz,mpts);add_fits<-cbind(add_fits,mtf)
}
else mtf<-add_fits[,addsz==mpts]
tmtf<-rep(0,nobs)
tmtf[mtf>=iseq[j]]<-1;tmtf[mtf<iseq[j]]<-0
mtfac<-sum(tmpf==y)/nobs
if(ltfac>utfac) {
  if(mtfac>=utfac) {upts<-mpts;utfac<-mtfac}
  else {upts<-lpts;utfac<-ltfac}
}
else if(ltfac<utfac) {
  if(mtfac>=ltfac) {lpts<-mpts;ltfac<-mtfac}
  else {lpts<-upts;ltfac<-utfac}
}
else{
  #cocok
  if(mtfac>=ltfac) {
    lpts<-mpts;ltfac<-mtfac
    upts<-mpts;utfac<-mtfac
  }
  else {upts<-lpts;utfac<-ltfac}
}
}
#true
if(ltfac>=utfac) {szfth[j]<-lpts;acfth[j]<-ltfac}
else {szfth[j]<-upts;acfth[j]<-utfac}
}

else if(nthc==nsrsz) {
  lpts<-szseq[nthc-1];upts<-szseq[nthc]
  ltfac<-acseq[nthc-1];utfac<-acseq[nthc]
  while(lpts!=upts) {
    mpts<-(lpts+upts)/2
    mpts<-floor(mpts)
    if(mpts%%2==0) mpts<-mpts+1
    if(mpts==upts) break
    if(length(addsz)==0) {
      mtf<-cv.fit(y,x,mpts,method)
      addsz<-c(addsz,mpts);add_fits<-cbind(add_fits,mtf)
    }
    else if(sum(addsz==mpts)==0) {
      mtf<-cv.fit(y,x,mpts,method)
      addsz<-c(addsz,mpts);add_fits<-cbind(add_fits,mtf)
    }
  }
}

```

```

else mtf<-add_fits[,addsz==mpts]
tmtf<-rep(0,nobs)
tmtf[mtf>=iseq[j]]<-1;tmtf[mtf<iseq[j]]<-0
mtfac<-sum(tmpf==y)/nobs
if(ltfac>utfac) {
  if(mtfac>=utfac) {upts<-mpts;utfac<-mtfac}
  else {upts<-lpts;utfac<-ltfac}
}
else if(ltfac<utfac) {
  if(mtfac>=ltfac) {lpts<-mpts;ltfac<-mtfac}
  else {lpts<-upts;ltfac<-utfac}
}
else{
  if(mtfac>=ltfac) {
    lpts<-mpts;ltfac<-mtfac
    upts<-mpts;utfac<-mtfac
  }
  else {upts<-lpts;utfac<-ltfac}
}
}
if(ltfac>=utfac) {szfth[j]<-lpts;acfth[j]<-ltfac}
else {szfth[j]<-upts;acfth[j]<-utfac}
}
else{
  lpts<-szseq[nthc-1];upts<-szseq[nthc]
  ltfac<-acseq[nthc-1];utfac<-acseq[nthc]
  while(lpts!=upts) {
    mpts<-(lpts+upts)/2
    mpts<-floor(mpts)
    if(mpts%%2==0) mpts<-mpts+1
    if(mpts==upts) break
    if(length(addsz)==0) {
      mtf<-cv.fit(y,x,mpts,method)
      addsz<-c(addsz,mpts);add_fits<-cbind(add_fits,mtf)
    }
    else if(sum(addsz==mpts)==0) {
      mtf<-cv.fit(y,x,mpts,method)
      addsz<-c(addsz,mpts);add_fits<-cbind(add_fits,mtf)
    }
  }
  else mtf<-add_fits[,addsz==mpts]
  tmtf<-rep(0,nobs)
  tmtf[mtf>=iseq[j]]<-1;tmtf[mtf<iseq[j]]<-0
  mtfac<-sum(tmpf==y)/nobs
  if(ltfac<utfac) {
    if(mtfac>=utfac) {upts<-mpts;utfac<-mtfac}
    else {upts<-lpts;utfac<-ltfac}
  }
  else if (ltfac<utfac){
    if(mtfac>=ltfac) {lpts<-mpts;ltfac<-mtfac}
    else {lpts<-upts;ltfac<-utfac}
  }
  else{
    if(mtfac>=ltfac) {

```

```

    lpts<-mpts;ltfac<-mtfac
    upts<-mpts;utfac<-mtfac
  }
  else { upts<-lpts;utfac<-ltfac }
}
if(ltfac>=utfac) { lsps<-lpts;lsbs<-ltfac }
else { lsps<-upts;lsbs<-utfac }
upts<-szseq[nthc+1];lpts<-szseq[nthc]
utfac<-acseq[nthc+1];ltfac<-acseq[nthc]
while(lpts!=upts) {
  mpts<-(lpts+upts)/2
  mpts<-floor(mpts)
  if(mpts%%2==0) mpts<-mpts+1
  if(mpts==upts) break
  if(length(addsz)==0) {
    mtf<-cv.fit(y,x,mpts,method)
    addsz<-c(addsz,mpts);add_fits<-cbind(add_fits,mtf)
  }
  else if(sum(addsz==mpts)==0) {
    mtf<-cv.fit(y,x,mpts,method)
    addsz<-c(addsz,mpts);add_fits<-cbind(add_fits,mtf)
  }
  else { mtf<-add_fits[,addsz==mpts] }
  tmtf<-rep(0,nobs)
  tmtf[mtf>=iseq[j]]<-1;tmtf[mtf<iseq[j]]<-0
  mtfac<-sum(tmtf==y)/nobs
  if(ltfac>utfac){
    if(mtfac>=utfac) { upts<-mpts;utfac<-mtfac }
    else { upts<-lpts;utfac<-ltfac }
  }
  else if(ltfac<utfac) {
    if(mtfac>=ltfac) { lpts<-mpts;ltfac<-mtfac }
    else { lpts<-upts;ltfac<-utfac }
  }
}
else{
  if(mtfac>=ltfac) {
    lpts<-mpts;ltfac<-mtfac
    upts<-mpts;utfac<-mtfac
  }
  else { upts<-lpts;utfac<-ltfac }
}
if(ltfac>=utfac) { usps<-lpts;usbs<-ltfac }
else { usps<-upts;usbs<-utfac }
if(lsbs>=usbs) { szfth[j]<-lsps;acfth[j]<-lsbs }
else { szfth[j]<-usps;acfth[j]<-usbs }
}
}
fnbst<-sum(max(acfth)==acfth);fscol<-seq(1:nbis)
if(fnbst==1){
  finsz<-szfth[max(acfth)==acfth]
}

```

```
    finth<-iseq[max(acfth)==acfth]
  }
  #true
  else {
    ftmpcol<-fscol[max(acfth)==acfth]
    tgcot<-ftmpcol[round(fnbst/2)]
    finsz<-szfth[tgcot]
    finth<-iseq[tgcot]
  }

  return(list(size=finsz,threshold=finth))
}

cv.fit <- function (y,x,npt,method) {
  num_pred<-ncol(x)
  num_obs<-nrow(x)
  lfit<-rep(0,num_obs)
  nv=3

  if(method=="lr") {
    lfit<-lr.cerp.cv(y,x,1,nv)$probability
  } else if(method=="lrt") {
    lfit<-lrt.cerp.cv(y,x,1,nv)$probability
  } else if(method=="ct") {lfit<-ct.cerp.cv(y,x,1,nv)$probability }
  return(lfit)
}
```

## RIWAYAT HIDUP



Muhamad Sabit Munawar, lahir di Kabupaten Cilacap pada 28 Desember 2001. Penulis merupakan anak keempat dari Sembilan bersaudara dari pasangan Bapak Sudirman Ibu Umiyah. Penulis telah menempuh Pendidikan mulai dari sekolah dasar di SDN 03 Gandrungmangu hingga kelas 4, dilanjutkan di SDN 020 Kuaro kelas 5 hingga lulus jenjang SD yang lulus pada 2013. Kemudian dilanjutkan pendidikan sekolah menengah atas di SMAN 01 Kuaro dan lulus pada tahun 2019. Pada tahun yang sama penulis melanjutkan Pendidikan di Universitas Islam Negeri Maulana Malik Ibrahim Malang Program Studi Matematika, Fakultas Sains dan Teknologi.

Selama menempuh Pendidikan tinggi, penulis aktif mengikuti beberapa kegiatan *volunteer*, organisasi baik didalam maupun diluar kampus. Organisasi dalam kampus yang pernah di ikuti adalah HMJ “Integral” Matematika UIN Maulana Malik Ibrahim Malang selama dua periode. Sedangkan organisasi luar kampus yang pernah di ikuti antara lain, Senyum Anak Nusantara Chapter Malang, IYOIN (*Indonesian Youth Opportunities in International Networking*) LC Malang dan beberapa organisasi lainnya. Selain itu, penulis juga pernah menjadi asisten praktikum Statistika Elementer. Penulis menerima segala saran dan kritik terhadap tugas akhir ini, untuk segala manfaatnya. Penulis dapat dihubungi melalui email: [muhamad.sabit.munawar@gmail.com](mailto:muhamad.sabit.munawar@gmail.com)



**KEMENTERIAN AGAMA RI  
UNIVERSITAS ISLAM NEGERI  
MAULANA MALIK IBRAHIM MALANG  
FAKULTAS SAINS DAN TEKNOLOGI**

Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

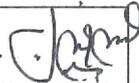
**BUKTI KONSULTASI SKRIPSI**

Nama : Muhamad Sabit Munawar  
NIM : 19610005  
Fakultas / Jurusan : Sains dan Teknologi / Matematika  
Judul Skripsi : Implementasi *Logistic Regression Ensemble* (LORENS)  
dalam Pengklasifikasian Angka Kematian Bayi di Indonesia  
Pembimbing I : Ria Dhea Layla N. K., M.Si.  
Pembimbing II : Mohammad Nafie Jauhari, M.Si.

No	Tanggal	Hal	Tanda Tangan
1.	13 Oktober 2022	Konsultasi Bab I	1.
2.	27 Oktober 2022	ACC Bab I	2.
3.	2 November 2022	Konsultasi Bab II dan III	3.
4.	10 Februari 2023	ACC Bab II dan III	4.
5.	10 Februari 2023	Konsultasi Kajian Agama Bab I dan II	5.
6.	14 Februari 2023	ACC Kajian Agama Bab I dan II	6.
7.	10 April 2023	ACC Seminar Proposal	7.
8.	9 Mei 2023	Konsultasi Revisi Seminar Proposal	8.
9.	10 Mei 2023	Konsultasi Bab IV dan V	9.
10.	17 Mei 2023	ACC Bab IV dan V	10.
11.	18 Mei 2023	Konsultasi Kajian Agama Bab IV	11.
12.	18 Mei 2023	ACC Kajian Agama Bab IV	12.
13.	26 Mei 2023	ACC Seminar Hasil	13.
14.	31 Mei 2023	Konsultasi Revisi Seminar Hasil	14.
15.	12 Juni 2023	ACC Sidang Skripsi	15.



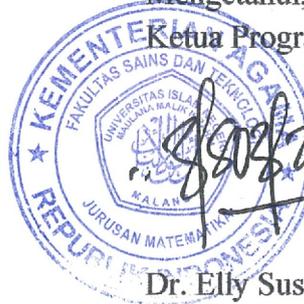
**KEMENTERIAN AGAMA RI**  
**UNIVERSITAS ISLAM NEGERI**  
**MAULANA MALIK IBRAHIM MALANG**  
**FAKULTAS SAINS DAN TEKNOLOGI**  
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

16.	22 Juni 2023	ACC Keseluruhan	16. 
-----	--------------	-----------------	---

Malang, 22 Juni 2023

Mengetahui,

Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc

NIP. 19741129 200012 2 005