

IMPLEMENTASI METODE *EXTREME GRADIENT BOOSTING* DALAM PENENTUAN KLASIFIKASI INDEKS KETIMPANGAN GENDER DI INDONESIA

SKRIPSI

**OLEH
HAFIZ ALDA CAHYAFITRI
NIM. 200601110116**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2025**

**IMPLEMENTASI METODE *EXTREME GRADIENT*
BOOSTING DALAM PENENTUAN KLASIFIKASI INDEKS
KETIMPANGAN GENDER 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
HAFIZ ALDA CAHYAFITRI
NIM. 200601110116**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2025**

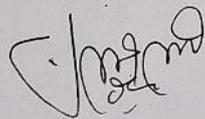
IMPLEMENTASI METODE *EXTREME GRADIENT BOOSTING* DALAM PENENTUAN KLASIFIKASI INDEKS KETIMPANGAN GENDER DI INDONESIA

SKRIPSI

Oleh
Hafiz Alda Cahyafitri
NIM. 200601110116

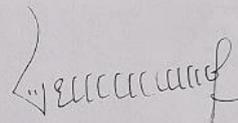
Telah Disetujui untuk Diuji
Malang, 2 Juni 2025

Dosen Pembimbing I



Ria Dhea Layla Nur Karisma, M.Si.
NIPPPK. 19900709 202321 2 037

Dosen Pembimbing II



Evawati Alisah, M.Pd.
NIP. 19720604 199903 2 001

Mengetahui,
Ketua Program Studi Matematika



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

IMPLEMENTASI METODE *EXTREME GRADIENT BOOSTING* DALAM PENENTUAN KLASIFIKASI INDEKS KETIMPANGAN GENDER DI INDONESIA

SKRIPSI

Oleh
Hafiz Alda Cahyafitri
NIM. 200601110116

Telah Dipertahankan di Depan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan untuk
Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal, 16 Juni 2025

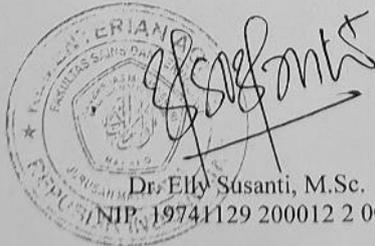
Ketua Penguji : Prof. Dr. Hj. Sri Harini, M.Si.

Anggota Penguji 1 : Dr. Fachrur Rozi, M.Si.

Anggota Penguji 2 : Ria Dhea Layla Nur Karisma, M.Si.

Anggota Penguji 3 : Evawati Alisah, M.Pd.

Mengetahui,
Ketua Program Studi Matematika


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

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Hafiz Alda Cahyafitri
NIM : 200601110116
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Implementasi Metode *Extreme Gradient Boosting* dalam Penentuan Klasifikasi Indeks Ketimpangan Gender di Indonesia

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

Malang, 16 Juni 2025

Yang membuat pernyataan,



Hafiz Alda Cahyafitri
NIM. 200601110116

MOTO

“Semua jatuh bangunmu hal yang biasa, angan dan pertanyaan waktu yang menjawabnya. Berikan tenggat waktu, bersedihlah secukupnya. Rayakan perasaanmu sebagai manusia”

(Baskara Putra)

PERSEMBAHAN

Dengan penuh rasa syukur yang tak terhingga, peneliti mengucapkan terima kasih kepada Allah SWT karena atas rahmat dan izin-Nya, peneliti berhasil menyelesaikan tahapan penulisan skripsi ini. Peneliti juga dengan tulus hati menyajikan skripsi ini untuk:

Bapak Kasiana dan Ibu Winarti, tiada kata yang dapat sepenuhnya menggambarkan rasa terima kasih atas doa, kasih sayang, usaha, dukungan, dan motivasi terbaik yang mereka berikan untuk kesuksesan peneliti. Kepada kakak peneliti, yang dengan tulus memberikan doa, nasihat, dan segala bentuk dukungan yang tak ternilai harganya.

Serta diri sendiri yang telah sabar untuk tetap bertahan dalam penyelesaian segala tanggung jawab dalam proses penulisan skripsi ini. Semoga persembahan ini dapat memberikan manfaat dan kebanggaan bagi semua yang terlibat.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur dipanjatkan kehadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga peneliti dapat menyelesaikan skripsi yang berjudul “Implementasi Metode *Extreme Gradient Boosting* dalam Penentuan Klasifikasi Indeks Ketimpangan Gender di Indonesia”. Sholawat serta salam senantiasa tercurahkan kepada junjungan Nabi Muhammad SAW yang menjadi suri tauladan sempurna dalam setiap aspek kehidupan.

Peneliti menyadari bahwa dalam penyusunan skripsi ini, peneliti membutuhkan bantuan serta bimbingan dari berbagai pihak. Sehingga dengan rendah hati, peneliti ingin menyampaikan rasa syukur dan mengucapkan terima kasih kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. 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 Nur Karisma, M.Si., selaku dosen pembimbing I yang telah melimpahkan pengetahuan, bimbingan, nasihat, motivasi, dan saran yang membangun dalam proses penyusunan penelitian ini.
5. Evawati Alisah, M.Pd., selaku dosen pembimbing II yang telah melimpahkan pengetahuan, bimbingan, nasihat, motivasi, dan saran yang membangun kepada peneliti.
6. Seluruh dosen Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang atas segala ilmu dan bimbingannya.
7. Seluruh keluarga terutama kedua orang tua dan kakak tersayang yang selalu memberikan do'a, motivasi, dan dukungan penuh dengan ikhlas secara moral dan materil.
8. Seluruh mahasiswa Program Studi Matematika angkatan 2020 yang telah

memberikan bantuan, semangat, dan motivasi sampai mencapai titik ini.

Semoga Allah SWT senantiasa memberikan balasan berupa rahmat dan karunia-Nya. Peneliti menyadari masih terdapat kekurangan dalam penyusunan laporan penelitian ini. Oleh karena itu, peneliti menerima saran dan kritik yang bersifat membangun dari semua pihak. Demikian skripsi ini disusun, peneliti berharap semoga laporan ini tidak hanya menjadi bagian dari tugas akhir, tetapi juga menjadi kontribusi kecil dalam upaya memperkaya ilmu pengetahuan bagi pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 16 Juni 2025

Peneliti

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN BERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
DAFTAR SIMBOL	xv
ABSTRAK	xvi
ABSTRACT	xvii
مستخلص البحث	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah.....	6
1.6 Definisi Istilah.....	6
BAB II KAJIAN TEORI	8
2.1 Klasifikasi	8
2.2 <i>Classification and Regression (CART)</i>	9
2.2.1 Pembentukan Pohon Klasifikasi	11
2.2.2 Pemangkasan Pohon Klasifikasi	15
2.2.3 Penentuan Pohon Klasifikasi Optimal	16
2.3 <i>Extreme Gradient Boosting</i>	17
2.3.1 Fungsi Objektif (<i>objective function</i>)	18
2.3.2 <i>Additive Training</i>	19
2.3.3 <i>Model Complexity</i>	20
2.4 Tingkat Akurasi Ketepatan Klasifikasi	23
2.5 Ketimpangan Gender	24
2.6 Indeks Ketimpangan Gender.....	25
2.7 Kajian Ketimpangan Gender dengan Teori Pendukung.....	27
2.8 Kajian Integrasi Ketimpangan Gender degan Al-Quran.....	29
BAB III METODE PENELITIAN	32
3.1 Jenis Penelitian.....	32
3.2 Data dan Sumber Data	32
3.3 Tahapan Penelitian	33
3.4 <i>Flowchart</i> Penelitian	35
BAB IV HASIL DAN PEMBAHASAN	36
4.1 Analisis Statistik Deskriptif	36

4.2	Analisis CART	45
4.2.1	Penentuan Data <i>Training</i> dan Data <i>Testing</i>	45
4.2.2	Pembentukan Pohon Klasifikasi CART	46
4.2.3	Tingkat Akurasi Ketepatan Klasifikasi CART	52
4.3	Analisis XGBoost	54
4.3.1	Pembentukan Pohon Klasifikasi <i>XGBoost</i>	57
4.3.2	Tingkat Akurasi Ketepatan Klasifikasi <i>XGBoost</i>	62
4.3.3	Variabel Important XGBoost	63
4.4	Kajian Keislaman dengan Hasil Penelitian	65
BAB V PENUTUP		67
5.1	Kesimpulan	67
5.2	Saran.....	68
DAFTAR PUSTAKA		69
LAMPIRAN.....		71
RIWAYAT HIDUP		74

DAFTAR TABEL

Tabel 2.1	<i>Confusion Matrix</i>	23
Tabel 3.1	Variabel Penelitian	32
Tabel 4.1	Analisis Deskriptif Persentase Persalinan Tidak Difasilitasi Kesehatan (<i>X1</i>).....	37
Tabel 4.2	Analisis Deskriptif Persentase Wanita Berusia 15-49 Tahun yang Pernah Kawin dan Saat Melahirkan Hidup Pertama Berusia Kurang dari 20 Tahun (<i>X2</i>).....	38
Tabel 4.3	Analisis Deskriptif Persentase Laki-laki yang Duduk di Parlemen (<i>X3</i>)	39
Tabel 4.4	Analisis Deskriptif Persentase Perempuan yang Duduk di Parlemen (<i>X4</i>)	40
Tabel 4.5	Analisis Deskriptif Persentase Penduduk Laki-laki dengan Pendidikan Minimal SMA (<i>X5</i>).....	41
Tabel 4.6	Analisis Deskriptif Persentase Penduduk Perempuan dengan Pendidikan Minimal SMA (<i>X6</i>).....	42
Tabel 4.7	Analisis Deskriptif Persentase Angkatan Kerja Laki-laki Terhadap Penduduk Usia Kerja (<i>X7</i>)	43
Tabel 4.8	Analisis Deskriptif Persentase Angkatan Kerja Perempuan Terhadap Penduduk usia Kerja (<i>X8</i>).....	44
Tabel 4.9	Pembagian Data <i>Training</i> dan Data <i>Testing</i>	46
Tabel 4.10	<i>Confusion Matrix</i> Klasifikasi CART.....	53
Tabel 4.11	Contoh Hasil Klasifikasi <i>XGBoost</i> IKG Kabupaten/Kota Tahun 2022.....	56
Tabel 4.12	Contoh Data IKG Perhitungan Manual <i>XGBoost</i>	57
Tabel 4.13	Perhitungan Kombinasi <i>Split</i> pada Fitur Persentase Persalinan Tidak Difasilitasi Kesehatan (<i>X1</i>).....	58
Tabel 4.14	Perhitungan <i>Gradient</i> dan <i>Hessian</i> pada Persentase Persalinan Tidak Difasilitasi Kesehatan (<i>X1</i>).....	59
Tabel 4.15	Perhitungan Kombinasi Nilai <i>Lambda</i>	60
Tabel 4.16	<i>Confusion Matrix XGBoost</i>	62

DAFTAR GAMBAR

Gambar 2.1 Pohon Classification and Regression Trees (CART)	10
Gambar 3.1 <i>Flowchart</i> Penelitian.....	35
Gambar 4.1 <i>Pie Chart</i> IKG 2022.....	36
Gambar 4.2 Hasil Pohon Klasifikasi pada IKG Kabupaten/Kota Tahun 2022	47
Gambar 4.3 Pohon Klasifikasi CART	48
Gambar 4.4 Pohon ke-1 <i>XGBoost</i>	54
Gambar 4.5 <i>Variable Important XGBoost</i>	64

DAFTAR LAMPIRAN

Lampiran 1	Data Indeks Ketimpangan Gender	71
Lampiran 2	<i>Syntax</i> Pembagian Data <i>Training</i> dan Data <i>Testing</i>	71
Lampiran 3	<i>Syntax</i> Pembentukan Pohon Klasifikasi <i>CART</i>	72
Lampiran 4	<i>Syntax</i> Pembentukan Pohon Klasifikasi <i>XGBoost</i>	72
Lampiran 5	<i>Syntax</i> Pembentukan Pohon Klasifikasi <i>XGBoost</i>	72
Lampiran 6	<i>Syntax</i> Evaluasi Model <i>XGBoost</i>	73
Lampiran 7	<i>Syntax</i> Plot <i>Variable Importance</i>	73

DAFTAR SIMBOL

Simbol-simbol yang digunakan dalam skripsi ini yaitu:

$i(t)$:	Fungsi heterogenitas indeks gini
$p(n t)$:	Proporsi kelas n pada node t
$p(m t)$:	Proporsi kelas m pada node t
$\phi(s,t)$:	Kriteria <i>goodness of split</i>
P_L	:	Peluang objek berada di node kiri
P_R	:	Peluang objek berada di node kanan
$i(t_L)$:	Nilai fungsi heterogenitas pada calon node kiri
$i(t_R)$:	Nilai fungsi heterogenitas pada calon node kanan
$N(t_L)$:	Banyak pengamatan pada calon node kiri
$N(t_R)$:	Banyak pengamatan pada calon node kanan
N	:	Total seluruh pengamatan
$p(m_0 t)$:	Proporsi kelas ke- m pada node t
$\hat{y}_i^{(t)}$:	<i>Final tree</i> model
$\hat{y}_i^{(t-1)}$:	Model pohon yang dihasilkan sebelumnya
$f_t(x_i)$:	Model baru yang dibangun
t	:	Jumlah total model dari <i>base tree models</i>
$l(y_i, \hat{y}_i^{(t)})$:	<i>Lost function</i>
$\Omega(f_i)$:	Regulasi
T	:	Jumlah <i>leaf</i> pada pohon ke- i
ω	:	Bobot <i>leaf</i> pada pohon ke- i

ABSTRAK

Cahyafitri, Hafiz Alda 2025. **Implementasi Metode *Extreme Gradient Boosting* dalam Penentuan Klasifikasi Indeks Ketimpangan Gender di Indonesia**. Skripsi. Progam Studi Matematika Fakultas Sains dan Tegnologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Evawati Alisah, M.Pd.

Kata kunci: Indeks Ketimpangan Gender, Klasifikasi, *Extreme Gradient Boosting*

Ketimpangan gender merupakan isu sosial yang penting di berbagai daerah, termasuk di Indonesia. Pengukuran dan klasifikasi indeks ketimpangan gender dapat membantu pemerintah dalam merumuskan kebijakan yang lebih efektif untuk mengatasi masalah ini. Penelitian ini bertujuan untuk mendiskripsikan implementasi metode *XGBoost* dan tingkat akurasi menggunakan metode *extreme gradient boosting* dalam penentuan klasifikasi indeks ketimpangan gender Kabupaten/Kota di Indonesia. Hasil klasifikasi menunjukkan terdapat 136 Kabupaten/Kota dengan kategori indeks ketimpangan gender rendah, 123 Kabupaten/Kota dengan skor indeks ketimpangan gender sedang, dan 204 Kabupaten/Kota dengan kategori indeks ketimpangan gender tinggi. Variabel dengan tingkat kepentingan paling berpengaruh adalah persalinan yang tidak dilakukan di fasilitas kesehatan (X_1). Hasil akurasi dari model klasifikasi sebesar 90,20% menunjukkan metode *extreme gradient boosting* mampu mengklasifikasikan Indeks ketimpangan gender Kabupaten/Kota di Indonesia dengan sangat baik.

ABSTRACT

Cahyafitri, Hafiz Alda. 2025. **Implementation of the Extreme Gradient Boosting Method in the Classification of Gender Inequality Index in Indonesia.** Undergraduate Thesis. Mathematics Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang. Advisors: (I) Ria Dhea Layla Nur Karisma, M.Si., (II) Evawati Alisah, M.Pd.

Keywords: Gender Inequality Index, Classification, Extreme Gradient Boosting

Gender inequality is an important social issue in many regions, including Indonesia. The measurement and classification of the gender inequality index can assist the government in formulating more effective policies to address this issue. This study aims to describe the implementation of the XGBoost method and the level of accuracy using the extreme gradient boosting method in determining the gender inequality index classification of districts/cities in Indonesia. The classification results show that there are 136 districts/cities with a low gender inequality index score, 123 districts/cities with a medium gender inequality index score, and 204 districts/cities with a high gender inequality index score. The variable with the most influential importance is childbirth that is not performed at a health facility (X_1). The accuracy of the classification model of 90,20% indicates that the extreme gradient boosting method is able to classify the gender inequality index of districts/cities in Indonesia very well.

مستخلص البحث

جهيا فطرى, حافظ ألدى (٢٠٢٥) تطبيق أسلوب التعزيز التدرجي الشديد في تصنيف مؤشر عدم المساواة بين الجنسين في إندونيسيا. البحث العلمي. قسم اللغة الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانخ. المشرفة: (١) ريا ديا ليلى نور كريسما، الماجستير في العلوم (٢) إيفاواتي أليسا، الماجستير في العلوم.

الكلمات الأساسية: مؤشر عدم المساواة بين الجنسين، التصنيف، التعزيز التدرجي الشديد

يعد عدم المساواة بين الجنسين قضية اجتماعية مهمة في مناطق مختلفة، بما في ذلك إندونيسيا. إن قياس وتصنيف مؤشرات عدم المساواة بين الجنسين من شأنه أن يساعد الحكومات على صياغة سياسات أكثر فعالية لمعالجة هذه المشكلة. هدفت هذه الدراسة إلى وصف تنفيذ طريقة *XGBoost* ومستوى الدقة باستخدام طريقة التعزيز المتدرج الشديد في تحديد تصنيف مؤشر عدم المساواة بين الجنسين في المناطق أو المدن في إندونيسيا. وظهرت نتائج التصنيف وجود ١٣٦ منطقة أو مدينة ذات فئة منخفضة من مؤشر عدم المساواة بين الجنسين، و ١٢٣ منطقة أو مدينة ذات درجة معتدلة من مؤشر عدم المساواة بين الجنسين، و ٢٠٤ منطقة أو مدينة ذات فئة عالية من مؤشر عدم المساواة بين الجنسين. المتغير ذو المستوى الأكثر تأثيراً من حيث الأهمية هو الولادة التي لا تتم في منشأة صحية (X_1). أشارت نتائج دقة نموذج التصنيف البالغة ٩٠,٢٠% إلى أن طريقة تعزيز التدرج الشديد قادرة على تصنيف مؤشر عدم المساواة بين الجنسين في المناطق أو المدن في إندونيسيا بشكل جيد للغاية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Klasifikasi merupakan metode statistika yang digunakan untuk mengelompokkan data secara sistematis berdasarkan karakteristik yang dimiliki (Lestawati dkk., 2018). Klasifikasi dilakukan untuk memudahkan pengambilan keputusan dalam suatu model dengan membagi data menjadi kategori atau kelas sesuai karakteristik tertentu. Terdapat beberapa pendekatan yang bisa diterapkan dalam klasifikasi, salah satunya merupakan metode *extreme gradient boosting* (*XGBoost*). *XGBoost* merupakan metode pembelajaran mesin yang menggunakan metode *ensemble* untuk klasifikasi dan regresi, *XGBoost* merupakan model dari algoritma *tree gradient boosting* dan memiliki kecepatan optimasi sepuluh kali lipat dibandingkan dengan metode *gradient boosting* lainnya (Chen & Guestrin, 2016).

Metode *XGBoost* memiliki beberapa kelebihan yang menjadikannya salah satu algoritma populer dalam *Machine Learning*. *XGBoost* menggunakan formulasi model yang lebih terstruktur untuk mengontrol *overfitting*, sehingga mendapatkan performa yang lebih baik (Punnoose & Ajit, 2016). Terlepas dari kelebihan metode *XGBoost* diakui sebagai model yang kuat, beberapa penelitian telah menemukan kelemahan dari metode *XGBoost*. Salah satunya *XGBoost* memiliki banyak parameter (*Hyperparameter*), yang mana dapat memengaruhi kinerja pada model. Namun, hal tersebut dapat diatasi dengan pemilihan parameter yang tepat untuk mencapai hasil akurasi optimal (Siringoringo dkk., 2021).

Penelitian menggunakan metode *XGBoost* dapat digunakan dalam berbagai jenis permasalahan, mulai dari kesehatan, ekonomi, sosial. Penelitian sebelumnya dilakukan oleh Haumahu dkk. (2021) dengan menerapkan metode *XGBoost* untuk mengklasifikasikan teks berita palsu di Indonesia. Hasil dari penelitian tersebut menunjukkan bahwa tingkat akurasi yang dihasilkan dengan metode *XGBoost* sebesar 92%. Selanjutnya penelitian menggunakan *XGBoost* juga dilakukan oleh Tingting Wang dkk. (2023) tentang pemantauan gempa regional dengan klasifikasi gempa tektonik magnitude rendah, ledakan, dan gempa akibat pertambangan. Pada penelitian dengan metode *XGBoost* diperoleh nilai akurasi 96,41% untuk gempa bumi, 90,38% untuk ledakan, dan 94,04% untuk gempa bumi yang disebabkan oleh pertambangan. Selanjutnya, penelitian lain tentang klasifikasi menggunakan algoritma *Machine Learning* pernah dilakukan oleh Anne Mudya Yolanda dkk. (2022). Penelitian tersebut mengusulkan klasifikasi untuk memprediksi Indeks Ketimpangan Gender (IKG) menggunakan metode *naïve bayes*, tingkat akurasi yang diperoleh pada keseluruhan data *training* sebesar 82,86% dan pada data *testing* sebesar 83,72% (Yolanda dkk., 2022). Hasil penelitian menunjukkan bahwa data IKG dapat diklasifikasi dengan baik menggunakan algoritma *Machine Learning*.

Penelitian ini menggunakan studi kasus Indeks Ketimpangan Gender (IKG) yang akan dianalisis dengan metode *XGBoost*. Pada tahun 2022, BPS mencatat ketimpangan gender Indonesia menempati peringkat ke-8 diantara negara-negara ASEAN. Berdasarkan hasil perhitungan, IKG Indonesia mengalami penurunan dari 0,4665 ditahun 2021 dan menjadi 0,459 pada tahun 2022. Pada tahun 2022, IKG tertinggi pada tingkat Kabupaten/Kota tercapai oleh Kabupaten Deiyai Provinsi

Papua Tengah dengan nilai 0,872, sementara IKG terendah tercatat di Kota Madiun Provinsi Jawa Timur sebesar 0,093 (Adi dkk., 2023). Capaian IKG di tingkat Kabupaten/Kota menunjukkan adanya disparitas ketimpangan gender di Indonesia. Hal ini menggambarkan kondisi pembangunan manusia di daerah-daerah tersebut belum tercapai, kerana masih adanya ketimpangan gender yang signifikan.

Ketimpangan gender merupakan suatu kondisi di mana terdapat perbedaan perlakuan, kesempatan, dan akses antara perempuan atau laki-laki. Penyebab ketimpangan gender sangat beragam dan berkaitan, salah satunya disebabkan oleh norma sosial dan budaya yang membedakan tingkahlaku antara perempuan dan laki-laki serta stereotip gender yang kaku, sehingga membatasi potensi dan kontribusi antara perempuan dan laki-laki (Sidiq & Erihadiana, 2022). Ketimpangan gender dapat terjadi dalam berbagai aspek kehidupan, termasuk pendidikan, pekerjaan, kesehatan, dan partisipasi politik. Salah satu isu ketimpangan gender yang disebabkan oleh stereotip kaku mengenai gender yaitu, *fatherless country* di Indonesia. Fenomena *fatherless country* menunjukkan adanya ketimpangan gender yang membatasi tingkah laku antara laki-laki dan perempuan dalam bermasyarakat, dimana hanya 28,3% peran ayah yang terlibat aktif dalam mengasuh anak di Indonesia (Adi dkk., 2023). Oleh karena itu, upaya untuk mengurangi ketimpangan gender sangat penting agar tercapai kesetaraan gender dengan memahami konsep gender secara mendalam.

Gender berbeda dengan jenis kelamin biologis, jenis kelamin secara biologis merupakan sebuah kodrat sebagaimana halnya manusia yang dilahirkan sebagai laki-laki atau perempuan. Sedangkan gender merupakan suatu konsep sosial berupa atribut maskulin atau feminim yang berasal dari kontruksi norma sosial dan

budaya. Atribut maskulin tidak harus dilekatkan pada jenis kelamin laki-laki dan sifat feminim juga tidak harus untuk perempuan, karena atribut-atribut tersebut bukan merupakan bawaan yang bersifat kodrati, melainkan terbentuk secara sosio-historis yang sifatnya tidak tetap, sehingga bisa dipertukarkan lintas seks (Apriliandra & Krisnani, 2020).

Pada Qur'an Kemenag (2022), ayat yang menunjukkan konsep kesetaraan antara laki-laki dan perempuan dalam Islam terdapat dalam Surah An-Nisa (4:32):

"Dan janganlah kamu iri hati terhadap apa yang telah Allah anugerahkan kepada sebagian kamu melebihi sebagian yang lain. Bagi laki-laki ada bahagian dari hasil usahanya, dan bagi perempuan (pun) ada bahagian dari hasil usahanya. Dan mohonlah kepada Allah sebagian dari karunia-Nya. Sesungguhnya Allah Maha Mengetahui segala sesuatu."

Ayat ini menegaskan bahwa baik laki-laki maupun perempuan memiliki bagian dan hak yang setara dalam hal usaha dan pemberian Allah. Ayat ini mengingatkan agar manusia tidak iri hati atas pemberian Allah terhadap orang lain, karena setiap individu, baik laki-laki maupun perempuan, akan memperoleh bagian dari hasil usahanya sendiri. Ini mencerminkan konsep kesetaraan gender dalam Islam, di mana baik laki-laki maupun perempuan memperoleh nilai dan hak yang setara di hadapan Allah (Ghoffar dkk., 2001).

Berdasarkan dasar latar belakang tersebut, maka dilakukan penelitian terkait klasifikasi pada Indeks Ketimpangan Gender Kabupaten/Kota di Indonesia menggunakan metode *XGBoost*. Penentuan variabel atau indikator yang berpengaruh terhadap Indeks Ketimpangan Gender berdasarkan pada penyempurnaan dari data SUSENAS dan SAKERNAS oleh BPS pada tahun 2021. Penelitian ini bertujuan untuk mendapat model terbaik dalam klasifikasi terhadap

Indeks Ketimpangan Gender Kabupaten/Kota di Indonesia menggunakan metode *XGBoost*.

1.2 Rumusan Masalah

Peneliti menetapkan rumusan masalah berdasarkan latar belakang yang telah dijabarkan, sebagai berikut:

1. Bagaimana implementasi metode *XGBoost* pada penentuan klasifikasi Indeks Ketimpangan Gender di Indonesia?
2. Bagaimana keakuratan *XGBoost* pada penentuan klasifikasi Indeks Ketimpangan Gender di Indonesia?

1.3 Tujuan Penelitian

Adapun tujuan berdasarkan latar belakang di atas, sebagai berikut:

1. Mendiskripsikan hasil implementasi *XGBoost* pada penentuan klasifikasi Indeks Ketimpangan Gender di Indonesia.
2. Menganalisis tingkat keakuratan *XGBoost* pada penentuan klasifikasi Indeks Ketimpangan Gender di Indonesia.

1.4 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian yang dilakukan berdasarkan tujuan penelitian sebelumnya:

1. Bagi Peneliti
Menambah wawasan mengenai implementasi metode *XGBoost* pada klasifikasi Indeks Ketimpangan Gender di Indonesia.

2. Bagi Program Studi

Sebagai bahan referensi dalam mengkaji serta menganalisis terkait pengklasifikasian data menggunakan metode *XGBoost*.

3. Bagi Instansi

Memberikan informasi terkait data Indeks Ketimpangan Gender di Indonesia, sehingga dapat digunakan sebagai gambaran maupun bahan pertimbangan dalam menentukan kebijakan selanjutnya.

1.5 Batasan Masalah

Batasan masalah pada penelitian ini sebagai berikut:

1. Data yang digunakan merupakan data (IKG) Indeks Ketimpangan Gender di Indonesia pada tahun 2022 dengan dimensi yang digunakan hanya mengacu pada laporan resmi IKG yang telah diterbitkan oleh badan pusat statistic (BPS), yaitu meliputi dimensi kesehatan, pemberdayaan dan pasar tenaga kerja.
2. Pengukuran tingkat akurasi pada klasifikasi menggunakan *confusion matrix* dan APER

1.6 Definisi Istilah

Klasifikasi : Proses identifikasi atau pengelompokan suatu objek ke dalam kategori tertentu berdasarkan karakteristik tertentu.

- Indeks : Ukuran atau nilai yang digunakan untuk merepresentasikan kondisi terhadap suatu standar.
- Ketimpangan Gender : Kondisi ketidaksetaraan antara laki-laki dan perempuan dalam berbagai aspek kehidupan.
- Machine Learning* : Ilmu pengembangan algoritma pada *system* komputer.
- XGBoost* : Metode *Machine Learning* dengan pohon keputusan, mengendalikan keputusan dan *overfitting*.
- Data *training* : Data yang digunakan untuk pembentukan model
- Data *testing* : Data yang digunakan untuk menguji kebaikan model

BAB II

KAJIAN TEORI

2.1 Klasifikasi

Klasifikasi merupakan metode dalam *Machine Learning*, yang merupakan bagian dari *artificial intelligence* (kecerdasan buatan) yang dapat melatih suatu sistem atau model untuk mempelajari data tanpa menuliskan algoritma secara eksplisit (Id, 2021). Terdapat tiga cabang utama dalam *Machine Learning*. Pertama adalah *Supervised Machine Learning*, di mana model dilatih menggunakan data yang telah diberi label untuk mempelajari pola dan mengkategorikan data baru ke dalam kategori yang tepat. Kedua, *Reinforcement Learning* di mana agen belajar dari interaksi dengan lingkungannya dan diberi imbalan untuk mengoptimalkan keputusan yang diambil, contohnya dalam permainan untuk memaksimalkan keuntungan. Ketiga, *Unsupervised Learning* yang bertujuan untuk mendeteksi pola atau struktur dalam data tanpa label, seperti pembentukan *cluster*. Algoritma-algoritma yang sering diterapkan dalam *Supervised Learning* ini misalnya *Decision Tree*, *Naïve Bayes*, *Support Vector Machine*, dan *Extreme Gradient Boosting (XGBoost)* (Han, 2011).

Machine Learning mempunyai dua pembelajaran utama yaitu, klasifikasi dan prediksi. Klasifikasi merupakan teknik analisis data yang bertujuan untuk mengidentifikasi model atau fungsi yang memisahkan suatu konsep atau kelas data berdasarkan atribut tertentu. Tujuan dari proses klasifikasi adalah untuk memprediksi kategori data yang tidak diketahui secara akurat dalam setiap permasalahan (Han, 2011).

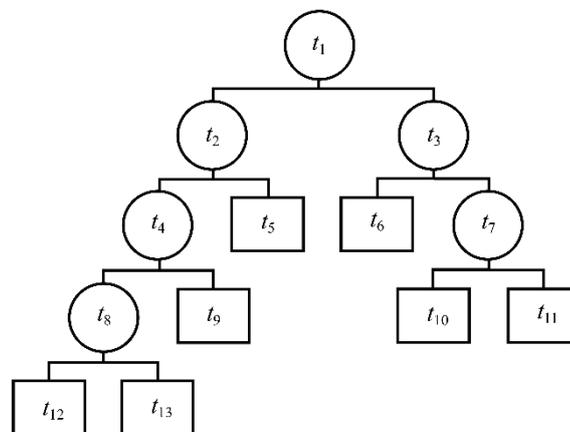
Terdapat langkah-langkah dalam proses klasifikasi yaitu (Kaur dkk., 2015):

1. Langkah pertama, model klasifikasi dibangun menggunakan data *training* dan data *testing*. Data *training* yang memiliki nilai label kelas yang diketahui digunakan untuk melatih algoritma klasifikasi, sedangkan data *testing* yang tidak memiliki label kelas digunakan untuk mengevaluasi performa model.
2. Langkah kedua, akurasi model klasifikasi dievaluasi menggunakan data *testing*. Jika hasil evaluasi menunjukkan performa yang memuaskan, model dapat digunakan untuk mengklasifikasikan data baru dengan label kelas yang tidak diketahui.

2.2 Classification and Regression (CART)

Classification and Regression Trees (CART) merupakan metode yang dikembangkan oleh Leo Breiman, Richard A. Olshen, Charles J. Stone, dan Jerome H. Friedman pada tahun 1984 yang menggunakan pendekatan pohon klasifikasi. Tujuan utama CART adalah untuk menghasilkan klasifikasi yang akurat, metode ini menggambarkan hubungan antara variabel dependen dan variabel independen dalam bentuk pohon klasifikasi. CART menggunakan metode klasifikasi pohon (*classification tree*) jika variabel dependen bersifat kategorik dan memiliki skala ordinal atau nominal. Sedangkan, jika variabel dependen bersifat kontinu dan memiliki skala interval atau rasio, maka metode regresi pohon (*regression tree*). Pohon klasifikasi pada metode CART sering disebut *binary decision tree* karena setiap *node* akan dibagi secara berulang menjadi dua *child node* secara berulang dan dimulai dari *parent node* (Breiman dkk., 1984).

Metode CART pohon klasifikasi sering disebut *binary decision tree* karena metode ini memisahkan *node* secara berulang-ulang menjadi dua *child node* dan dimulai dengan *parent node*. *Parent node* merupakan simpul awal dari pohon yang mewakili seluruh data sebelum dipisahkan menjadi kelas-kelas berdasarkan karakteristik tertentu (Breiman dkk., 1984). Berikut merupakan ilustrasi pohon klasifikasi yang dimulai dari *parent node* hingga mencapai *terminal node*.



Gambar 2.1 Pohon *Classification and Regression Trees (CART)*

Proses pembentukan pohon keputusan pada metode CART disebut sebagai pemilihan rekursif biner (*binary recursive partitioning*). Dalam proses ini data dipisahkan secara berulang-ulang untuk membentuk banyak node hingga membentuk pohon klasifikasi. Proses ini melibatkan pemisahan data secara berulang-ulang hingga data tidak dapat lagi dipisah dan mencapai *terminal node*. Setiap *parent node* akan dibagi menjadi dua *child node* yang lebih spesifik dengan tujuan memaksimalkan pemisahan data yang jelas antar kelas dalam dataset. Konsep *partitioning* juga dikenal dalam metode CART sebagai langkah klasifikasi untuk mendapatkan sampel dengan ukuran yang lebih kecil (Lewis, 2000). Pada

metode CART, simpul awal disebut sebagai *parent node* notasinya adalah t_1 . Selanjutnya *parent node* menghasilkan *child node* yang dinotasikan dengan t_2, t_3, t_4, t_7, t_8 . Hasil akhir pembentukan pohon klasifikasi disebut *terminal node* yang dinotasikan dengan $t_5, t_6, t_9, t_{10}, t_{11}, t_{12}, t_{13}$, pada penelitian ini berbentuk persegi. Berikut adalah ilustrasi struktur dari pohon klasifikasi.

2.2.1 Pembentukan Pohon Klasifikasi

Proses pembentukan model pohon klasifikasi diawali dengan menentukan variabel yang digunakan sebagai pemisah data. Nilai dari variabel tersebut, yang dikenal sebagai nilai *threshold* digunakan sebagai pemilah pada setiap *node*. Dalam membangun model pohon klasifikasi diperlukan data *training* yang terdiri dari sampel L dengan N pengamatan. Proses pembentukan pohon klasifikasi melibatkan tiga tahapan, yaitu sebagai berikut (Breiman dkk., 1984):

1. Penentuan Pemilah (*Splitting Rule*)

Proses penentuan pemilah bertujuan untuk mengurangi tingkat keheterogenan pada *parent node* dan meningkatkan homogenitas data pada *child node*. Proses ini menggunakan data *training* sebanyak L sampel dan menerapkan kriteria *goodness of split*. Tingkat heterogen suatu kelas dalam pohon klasifikasi diukur melalui *node* dengan nilai *impurity*, fungsi *impurity* yang umumnya dihitung menggunakan indeks gini. Indeks gini memiliki beberapa keunggulan seperti dapat digunakan dalam berbagai masalah, mudah dipahami, dan perhitungan yang sederhana (Breiman,

dkk., 1984). Secara umum persamaan indeks gini dapat dituliskan sebagai berikut:

$$i(t) = \sum_{n=1, n \neq m}^k p(n|t) p(m|t) \quad (2.1)$$

dimana:

$i(t)$: Fungsi heterogenitas indeks gini

$p(n|t)$: Proporsi kelas n pada *node t*

$p(m|t)$: Proporsi kelas m pada *node t*

Sehingga, pada *node t* kelas ke- m indeks gini dapat didefinisikan dengan:

$$\begin{aligned} i(t) &= \sum_{n=1, n \neq m}^k p(n|t) p(m|t) \\ &= \sum_{m=1}^k p(m|t) - p^2(m|t) \\ &= \sum_{m=1}^k p(m|t) - \sum_{m=1}^k p^2(m|t) \\ &= 1 - \sum_{m=1}^k p^2(m|t) \end{aligned} \quad (2.2)$$

Penentuan pemilah terbaik menggunakan nilai *goodness of split*, pemilah terbaik yang memiliki nilai *goodness of split* terbesar. *Goodness of split* dapat diartikan sebagai penurun keheterogenan, dengan persamaan sebagai berikut:

$$\begin{aligned} \phi(s, t) &= \Delta i(s, t) \\ &= i(t) - P_L \cdot i(t_L) - P_R \cdot i(t_R) \end{aligned} \quad (2.3)$$

dimana:

$\phi(s, t)$: Kriteria *goodness of split*

P_L : Peluang objek berada di *node* kiri

P_R : Peluang objek berada di *node* kanan

$i(t_L)$: Nilai fungsi heterogenitas pada calon *node* kiri

$i(t_R)$: Nilai fungsi heterogenitas pada calon *node* kanan

Untuk menentukan peluang objek pada *node* kiri dan kanan adalah sebagai berikut:

$$P_L = \frac{N(t_L)}{N} \quad (2.4)$$

$$P_R = \frac{N(t_R)}{N} \quad (2.5)$$

dimana:

$N(t_L)$: Banyak pengamatan pada calon *node* kiri

$N(t_R)$: Banyak pengamatan pada calon *node* kanan

N : Total seluruh pengamatan

2. Penentuan Terminal Node

CART menggunakan beberapa kriteria untuk menentukan apakah sebuah *node* harus diubah menjadi *terminal node* atau tidak. Jika penurunan keheterogenan pada *node* t tidak signifikan, maka *node* t tidak akan diubah menjadi *terminal node*. Penurunan keheterogenan diukur menggunakan standar *goodness of split* pada indeks gini. Jika hanya ada satu pengamatan pada *node* t atau jika jumlah kasus minimum pada *node* t kurang dari atau

sama dengan lima ($n \leq 5$), akan diubah menjadi *terminal node*. Pohon klasifikasi CART akan berhenti ketika tingkat kedalaman pohon telah tercapai.

3. Penandaan Label Kelas

Penandaan label kelas pada terminal *node* bertujuan memperoleh informasi karakteristik dan klasifikasi dari hasil pengamatan pada setiap kelas variabel dependen yang terbentuk. Aturan jumlah terbanyak digunakan untuk menetapkan pelabelan kelas pada terminal *node*, pada penandaan label kelas diilustrasikan pada persamaan:

$$p(m_0|t) = \max p(m|t) = \max \frac{N_m(t)}{N(t)} \quad (2.6)$$

dimana:

$p(m_0|t)$: Proporsi kelas ke- m pada *node* t

$N_m(t)$: Banyaknya kelas ke- m yang diamati terhadap terminal *node* t

$N(t)$: Banyaknya pengamatan pada terminal *node*

m_0 : Label kelas pada terminal *node* t

Pada proses ini label kelas m_0 pada terminal *node* t berdasarkan asumsi bahwa setiap kelas memiliki tingkat kesalahann klasifikasi yang sama.

Dengan demikian, penetapan label kelas m_0 akan menghasilkan estimasi kesalahann klasifikasi minimal pada terminal *node* t sebesar

$$r(t) = 1 - p(m_0|t)$$

2.2.2 Pemangkasan Pohon Klasifikasi

Pemangkasan pohon (*Pruning*) dilakukan untuk mendapatkan pohon klasifikasi yang optimal. *Pruning* merupakan proses pengurangan *node* pohon dengan mempertimbangkan ukuran pohon tanpa mengurangi akurasi prediksi. Proses *Pruning* menggunakan metode *cost complexity*, rumus ukuran *cost complexity* yang memiliki nilai $a > 0$ dapat menggunakan persamaan sebagai berikut (Breiman dkk., 1984):

$$R_\alpha(T) = R(T) + a[\tilde{T}] \quad (2.7)$$

dimana:

- $R_\alpha(T)$: *Resubstitution* suatu pohon T terhadap kompleksitas a
- $R(T)$: Pendugaan pengganti (*resubstitution estimate*)
- α : Parameter *cost complexity* untuk penambahan satu *node* pohon T
- \tilde{T} : Banyak *node* pada terminal pohon T

Tahapan dalam proses pemangkasan pohon klasifikasi melibatkan pemilihan t_R dan t_L dari T_{\max} yang dihasilkan dari *parent node*. Jika *child node* dan *parent node* memenuhi persamaan $R(t) = R(t_R) + R(t_L)$, maka proses pemangkasan dapat dilakukan. Dari proses ini akan didapatkan pohon T_1 yang memenuhi persamaan $R(t_1) = R(t_{\max})$. Pohon klasifikasi yang optimal diperoleh dengan mengulangi proses ini secara berulang hingga tidak ada lagi pemangkasan yang memungkinkan.

2.2.3 Penentuan Pohon Klasifikasi Optimal

Pohon klasifikasi yang optimal dapat ditentukan dari ukuran pohon, di mana pohon yang terlalu besar dan mengalami *overfitting* dapat menyebabkan nilai *cost complexity* meningkat. Hal tersebut terjadi karena struktur data yang terlalu kompleks, sehingga diperlukan pemilihan pemilah pengganti terkecil dari pohon optimal. Jika pengganti terkecil dilambangkan dengan $R(t)$, maka dipilih pohon terbesar dikarenakan nilai $R(t)$ yang kecil menghasilkan pohon yang besar.

Pemilihan pohon klasifikasi yang optimal menggunakan metode *test sample estimate*. Tahapan ini melibatkan pembagian data menjadi dua bagian, yaitu data *training* dilambangkan dengan $L1$ dan data *testing* dilambangkan dengan $L2$. Penentuan pohon klasifikasi menggunakan data *training* dan penentuan nilai $R^{ts}(T_t)$ menggunakan data *testing*. Persamaan *test sample estimate* sebagai berikut:

$$R^{ts}(T_t) = \frac{1}{N_2} \sum_{(x_n, m_n) \in L_2} X(d(x_n) \neq m_n) \quad (2.8)$$

dimana:

$R^{ts}(T_t)$: Total proporsi kesalahan pada *test sample estimate*

N_2 : Banyak pengamatan pada data *training*

$X(d(x_n) \neq m_n)$: Hasil klasifikasi

Pemilihan pohon klasifikasi yang optimal bertujuan untuk mengestimasi proporsi kesalahan pada pohon klasifikasi, sehingga dapat ditentukan pohon klasifikasi optimal T_t dengan $R^{ts}(T_t) = \min_t R^{ts}(T_t)$.

2.3 *Extreme Gradient Boosting*

Extreme gradient boosting atau *XGBoost* merupakan salah satu metode dalam *machine learning* yang dikembangkan oleh (Chen & Guestrin, 2016). *XGBoost* merupakan salah satu metode yang bisa melakukan prediksi maupun klasifikasi. Metode ini menggunakan struktur pohon keputusan sebagai dasarnya. *XGBoost* merupakan salah satu metode boosting yang terdiri dari beberapa pohon keputusan, di mana nilai setiap pohon akan diperbarui secara bertahap sehingga menghasilkan pohon klasifikasi yang kuat dan akurat.

Secara sistematis proses pembentukan prediksi model menggunakan metode *XGBoost* dapat dinyatakan dalam bentuk persamaan sebagai berikut:

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i)$$

$f_k(x_i)$ mempresentasikan model pohon keputusan ke-k yang digunakan untuk memprediksi nilai target. Untuk y_i didapatkan dari perhitungan berikut:

$$\hat{y}_i^{(0)} = 0$$

$$\hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i)$$

$$\hat{y}_i^{(2)} = f_1(x_1) + f_2(x_2) = \hat{y}_i^{(1)} + f_2(x_2)$$

⋮

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_i(x_i) \tag{2.9}$$

dimana:

$\hat{y}_i^{(t)}$: *Final tree model*

$\hat{y}_i^{(t-1)}$: Model pohon yang dihasilkan sebelumnya

- $f_t(x_i)$: Model baru yang dibangun
 t : Jumlah total model dari *base tree models*

2.3.1 Fungsi Objektif (*objective function*)

Fungsi objektif dalam *XGBoost* merupakan sebuah fungsi yang mengukur seberapa baik model dalam mempelajari pola sesuai dengan data *training*. Fungsi ini menggabungkan fungsi kerugian (*loss function*) yang menghitung kesalahan prediksi dilambangkan dengan $L(\theta)$, dan *regulation term* dilambangkan sebagai $\Omega(\theta)$ berfungsi membantu mencegah *overfitting* dengan mengendalikan kompleksitas model. Gabungan kedua komponen ini membentuk fungsi objektif secara keseluruhan, yang secara matematis dapat dinyatakan sebagai:

$$\begin{aligned} obj(\theta) &= L(\theta) + \Omega(\theta) \\ obj(\theta) &= \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^t \Omega(f_k) \end{aligned} \tag{2.10}$$

dimana:

$obj(\theta)$: *objective function*

$L(\theta)$: *loss function*

$\Omega(\theta)$: *regulation term*

Pada masalah klasifikasi multikelas, *loss function* yang akan digunakan adalah fungsi *cross entropy loss*. Fungsi ini dirancang untuk memperbesar kemungkinan prediksi yang benar dan memperkecil kemungkinan prediksi yang salah. Bentuk matematis dari *loss function* sebagai berikut:

$$L(\theta) = -[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (2.11)$$

2.3.2 Additive Training

Additive training merupakan pendekatan dimana model diditingkatkan secara bertahap dengan menambahkan pohon keputusan baru pada setiap iterasi, dengan tujuan memperbaiki kesalahan prediksi yang dibuat oleh pohon sebelumnya.

Persamaan *objective function* pada *additive training* sebagai berikut:

$$obj(t) = \sum_{i=1}^n -[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] + \sum_{i=1}^t \Omega(f_i) \quad (2.12)$$

Pada kasus umum, untuk mengoptimalkan *objective function* dilakukan dengan memanfaatkan pendekatan *taylor expansion* hingga *second order* pada *loss function*. Proses ini melibatkan informasi *gradient* sebagai turunan pertama dari *loss function* dan *hessian* sebagai turunan keduanya. Dengan menggunakan informasi ini dapat melakukan pembaruan model yang lebih efisien dan akurat. Sehingga diperoleh *objective function* sebagai berikut:

$$obj(t) = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_i) + C \quad (2.13)$$

dengan g_i dan h_i didefinisikan sebagai persamaan berikut:

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \quad (2.14)$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) \quad (2.15)$$

Bentuk sederhana *objective function* dengan menghapus konstanta:

$$obj(t) = \sum_{i=1}^n \left[g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_i) \quad (2.16)$$

2.3.3 Model Complexity

Model complexity mengacu pada pada tingkat kerumitan model yang dibangun. Semakin kompleks model semakin besar juga kemungkinan model tersebut mengalami *overfitting*. Oleh karena itu kompleksitas model harus dikendalikan agar model dapat memperbaiki pohon baru dengan baik. Kompleksitas model diukur menggunakan *regulation term* yang didefinisikan dalam persamaan sebagai berikut:

$$f_t(x) = w_{q(x)}, w \in R^t, q: R^d \rightarrow \{1, 2, \dots, t\} \quad (2.17)$$

Dimana w adalah *vector* skor dari *leaf*, q adalah fungsi yang menempatkan setiap input data ke *leaf* yang sesuai, t adalah jumlah *leaf*. *Regulation term* merupakan fungsi penalti untuk mencegah model terlalu kompleks yang dapat menyebabkan *overfitting*. *Overfitting* terjadi ketika model terlalu menyesuaikan diri terhadap data *training* sehingga performanya menurun saat diuji pada data baru. Oleh karena itu, *regulation term* berfungsi sebagai pengontrol agar model tetap sederhana namun tetap memiliki kemampuan yang akurat. Persamaan *regulation term* sebagai berikut:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (2.18)$$

dimana:

$\Omega(f)$: *regulation term*

γ : Paramater kompleksitas model

T : Jumlah *leaf* dalam sebuah pohon keputusan

λ : Parameter regulasi untuk penalty bobot *leaf*

w : Skor prediksi pada *leaf*

didapat persamaan *objective function* sebagai berikut:

$$\begin{aligned} obj^{(t)} &\approx \sum_{i=1}^n \left[g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \\ &= \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \end{aligned} \quad (2.19)$$

Fungsi objektif *XGBoost* yang sebelumnya telah dirumuskan dapat disederhanakan lebih lanjut untuk mempermudah proses optimasi pada saat pembentukan pohon keputusan. Penyederhanaan ini menghasilkan persamaan sebagai berikut:

$$obj^{(t)} = \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \quad (2.20)$$

dimana:

G_j : Jumlah *gradient* dari seluruh data pada *leaf*

H_j : Jumlah *hessian* dari seluruh data pada *leaf*

λ : Parameter regulasi untuk penalty bobot *leaf*

γ : Paramater kompleksitas model

T : Jumlah *leaf* dalam sebuah pohon keputusan

Optimasi struktur pohon keputusan tidak dilakukan sekaligus, melainkan dilakukan bertahap setiap tingkat percabangan (*split*) dari satu *node* ke dua *leaf*. Tujuan dari proses ini adalah untuk memutuskan apakah suatu *node* layak dipecah berdasarkan nilai *gain*, yaitu skor yang menggambarkan peningkatan fungsi

objektif akibat pemecahan tersebut. Nilai *gain* dihitung menggunakan persamaan berikut:

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (2.21)$$

dimana:

- G_L dan G_L : Jumlah *gradient* dan *hessian* pada *leaf* kiri
- G_R dan G_R : Jumlah *gradient* dan *hessian* pada *leaf* kanan
- λ : Parameter regulasi untuk *penalty* bobot *leaf*
- γ : *Threshold* nilai *gain* yang diperlukan pada *split*

Parameter λ berfungsi sebagai regulasi L2 yang mengendalikan besar kecilnya bobot *leaf* untuk menghindari *overfitting*. Pemilihan nilai λ secara umum dimulai dari 0 hingga 10, dan dapat disesuaikan melalui proses tuning berdasarkan kompleksitas data dan kebutuhan generalisasi model. Sementara itu, nilai γ merupakan *threshold* tetap (*pruning parameter*) yang digunakan untuk menentukan apakah suatu pemisahan layak dilakukan. *Split* hanya akan dilakukan jika nilai *gain* melebihi nilai γ , artinya pemisahan tersebut memberikan peningkatan yang signifikan terhadap fungsi objektif. Jika nilai *gain* lebih kecil dari γ atau bahkan negatif, maka *split* tidak dilakukan dan *node* tersebut akan dianggap sebagai *leaf*. *Split* hanya akan ditambahkan jika menghasilkan *gain* yang lebih besar dari γ , nilai λ dan γ digunakan untuk mengendalikan kompleksitas struktur pohon keputusan serta menjaga keseimbangan antara akurasi dan efisiensi model (Chen & Guestrin, 2016).

2.4 Tingkat Akurasi Ketepatan Klasifikasi

Menghitung ukuran ketepatan klasifikasi dalam analisis pengelompokan dilakukan dengan tujuan mengetahui kebaikan model dalam klasifikasi data yang telah dilakukan. Terdapat banyak cara untuk menghitung ketepatan klasifikasi, salah satunya menggunakan *Apparent Error Rate* (APER). *Apparent Error Rate* (APER) merupakan pendekatan untuk mengetahui tingkat akurasi hasil klasifikasi dengan melihat nilai eror dari suatu klasifikasi (Prasetyo, 2012). Untuk menghitung nilai APER diperlukan tabel perbandingan antara hasil klasifikasi berdasarkan observasi dan hasil klasifikasi berdasarkan prediksi suatu metode yang disebut dengan *confusion matrix*. *Confusion matrix* merupakan metode yang digunakan dalam analisis data klasifikasi untuk mengevaluasi hasil dari model yang telah dianalisis. Berikut ini merupakan tabel metode *confusion matrix* untuk masalah klasifikasi multikelas (khotimah & Utami, 2022).

Tabel 2.1 *Confusion Matrix*

<i>Confusion Matrix</i>		Kelas Aktual			
		C_1	C_2	...	C_N
Kelas Prediksi	C_1	$C_{1,1}$	FN	...	$C_{1,N}$
	C_2	FP	TP	...	FP

	C_N	$C_{N,1}$	FN	...	$C_{N,N}$

Pada Tabel 2.1 terdapat empat istilah dari istilah perhitungan *confusion matrix*. Pertama, *true positive* (TP) yaitu hasil prediksi model kelas positif tepat sesuai dengan data aktual. Kedua, *false positive* (FP) yaitu menunjukkan hasil prediksi model kelas negatif yang berada di kelas positif. Kemudian *false negative* (FN) yang menunjukkan hasil dari prediksi model kelas positif yang berada di kelas

negatif, dan yang terakhir adalah *true negative* (TN) yang berarti hasil prediksi model kelas negatif tepat sesuai data aktual yang ada. Suatu model dapat dikatakan memiliki tingkat akurasi yang baik jika mempunyai nilai APER yang kecil dan tingkat akurasi yang tinggi, perhitungan nilai APER dengan persamaan sebagai berikut (Hilmiyah, 2017):

$$APER = \left(\frac{FP + FN}{TP + TN + FP + FN} \right) \times 100$$

Dan persamaan tingkat akurasi sebagai berikut:

$$\text{Tingkat akurasi} = 1 - APER$$

2.5 Ketimpangan Gender

Gender merupakan konsep sosial yang terbentuk karena norma dan budaya masyarakat dalam membedakan peran, perilaku, dan tanggung jawab antara laki-laki dan perempuan. Gender berbeda dengan seks yang ditentukan secara biologis, gender bersifat dinamis yang dapat berubah seiring waktu dan budaya setempat (Apriliandra & Krisnani, 2020). Norma-norma sosial yang kaku dan stereotip negatif tentang gender seringkali membatasi peran, perilaku, dan tanggung jawab antara laki-laki dan perempuan, hal tersebut memicu terjadinya ketimpangan gender dalam masyarakat.

Ketimpangan gender merupakan kondisi dimana laki-laki atau perempuan dirugikan karena terdapat perbedaan hak, status, dan perlakuan antara laki-laki dan perempuan (Sidiq & Erihadiana, 2022). Ketimpangan gender dapat terjadi di berbagai aspek kehidupan, termasuk pendidikan, pekerjaan, politik, dan kesehatan. Ketimpangan gender dalam aspek pendidikan, perempuan kesulitan mendapatkan akses pendidikan tinggi karena norma budaya dan pernikahan dini. Ketimpangan

gender juga terjadi dalam aspek ketenagakerjaan, tercermin dalam upah yang tidak setara antara laki-laki dan perempuan walaupun berada pada tingkat yang sama. Selain itu, ketimpangan gender dalam aspek politik terlihat dari rendahnya keterwakilan perempuan dalam lembaga politik dan pengambilan putusan. Ketimpangan gender juga terjadi pada aspek kesehatan, dimana perempuan sering mendapatkan keterbatasan layanan kesehatan terutama dalam hal kesehatan reproduksi (Apriliandra & Krisnani, 2021).

2.6 Indeks Ketimpangan Gender

Indeks ketimpangan gender (IKG) merupakan alat ukur untuk mengetahui tingkat ketimpangan gender dalam suatu wilayah agar dapat dijadikan evaluasi terhadap program penangan yang dilakukan. Pada tahun 2017, BPS mengembangkan kajian pengukuran IKG yang merujuk pada *Gender Inequality Index* (GII) oleh UNDP. Indikator IKG dikelompokan dalam tiga dimensi, yaitu kesehatan, pemberdayaan dan pasar tenaga kerja. Dimensi kesehatan diukur dari persentase persalinan tidak difasilitasi kesehatan dan persentase wanita berusia 15-49 tahun yang pernah kawin dan saat melahirkan hidup pertama berusia kurang dari 20 tahun. Dimensi pemberdayaan diukur dari persentase penduduk laki-laki dan perempuan dengan jenjang pendidikan minimal SMA dan persentase laki-laki dan perempuan yang duduk di parlemen. Dimensi pasar kerja diukur dari tingkat partisipasi angkatan kerja (TPAK) laki-laki dan perempuan. Nilai IKG bervariasi antara 0 hingga 1, dimana nilai 0 menggambarkan adanya kesetaraan antara laki-laki dan perempuan, sementara nilai 1 menggambarkan adanya ketimpangan antara laki-laki dan perempuan dalam semua aspek. Nilai IKG diperoleh melalui beberapa

tahapan perhitungan indeks, yaitu menghitung Indeks Perempuan dan Indeks laki-laki (Adi dkk., 2023).

Indeks perempuan dihitung dengan persamaan sebagai berikut:

$$G_F = \sqrt[3]{\left(\frac{0,01}{Faskes} \cdot \frac{0,01}{ULP}\right)^{\frac{1}{2}} \cdot (PR_F \cdot SE_F)^{\frac{1}{2}} \cdot TPAK_F} \quad (2.22)$$

Indeks Laki-laki dihitung dengan persamaan sebagai berikut:

$$G_M = \sqrt[3]{1 \cdot (PR_M \cdot SE_M)^{\frac{1}{2}} \cdot TPAK_M} \quad (2.23)$$

Kemudian perlu menghitung gabungan indeks perempuan dan indeks laki-laki dengan rata-rata setara untuk menghasilkan indeks yang terdistribusi setara. Penjumlahan dengan rata-rata setara digunakan untuk menggambarkan ketimpangan gender berdasarkan hubungan antara indikator dalam IKG. Penjumlahan indeks laki-laki dan indeks perempuan dapat diperoleh dari:

$$HARM(G_F, G_M) = \left[\frac{(G_F)^{-1} + (G_M)^{-1}}{2} \right]^{-1} \quad (2.24)$$

Selanjutnya mengukur indeks ketimpangan gender menggunakan rumus:

$$IKG = 1 - \frac{HARM(G_F, G_M)}{G_{F,M}} \quad (2.25)$$

Keterangan:

G_F : Indeks Perempuan

G_M : Indeks Laki-laki

$Faskes$: Pesentase persalinan tidak di fasilitasi kesehatan

ULP	: Persentase wanita berusia 15-49 tahun yang pernah kawin dan saat melahirkan hidup pertama berusia kurang dari 20 tahun
PR_F	: Persentase perempuan yang duduk di parlemen
PR_M	: Persentase laki-laki yang duduk di parlemen
SE_F	: Persentase perempuan dengan pendidikan minimal SMA
SE_M	: Persentase laki-laki dengan pendidikan minimal SMA
$TPAK_F$: Persentase angkatan kerja perempuan terhadap penduduk perempuan usia kerja
$TPAK_M$: Persentase angkatan kerja laki-laki terhadap penduduk laki-laki usia kerja terhadap penduduk laki-laki usia kerja
$HARM(G_F, G_M)$: Gabungan Indeks laki-laki dan Indeks Perempuan

2.7 Kajian Ketimpangan Gender dengan Teori Pendukung

Ketimpangan gender merupakan permasalahan sosial yang dapat dialami pada laki-laki maupun perempuan dan seringkali disebabkan oleh ketidaksetaraan hak dalam bermasyarakat. Pemerintah berupaya mengurangi ketimpangan gender di Indonesia melalui perhitungan Indeks Ketimpangan Gender (IKG) yang terdiri dari delapan indikator. Indikator-indikator ini akan dikelompokkan untuk menghasilkan klasifikasi yang dapat menentukan tingkat ketimpangan gender di berbagai wilayah Indonesia, apakah rendah, sedang, atau tinggi. Oleh karena itu, salah satu metode yang dapat digunakan untuk melakukan klasifikasi pada data IKG

adalah metode *XGBoost*. Metode ini dapat meningkatkan akurasi prediksi dengan menggunakan pohon keputusan yang sesuai dengan variabel-variabel yang berpengaruh. Langkah pertama adalah melakukan pengumpulan data. Langkah kedua adalah melakukan analisis deskriptif yang bertujuan untuk mendapatkan gambaran umum mengenai karakteristik data pada penelitian, yaitu Indeks Ketimpangan Gender pada 514 kabupaten/kota di Indonesia. Langkah ketiga adalah melakukan pemisahan data menjadi data *training* dan data *testing*. Kemudian, model klasifikasi dibentuk dengan metode *XGBoost* menggunakan data *training*. Tahap awal dalam membentuk model klasifikasi adalah melatih model secara bertahap, di mana setiap model baru memperbaiki kesalahan model sebelumnya.

Model klasifikasi *XGBoost* yang telah terbentuk akan memberikan klasifikasi Indeks Ketimpangan Gender pada 513 kabupaten/kota di Indonesia ke dalam tiga kelas. Jika hasil *XGBoost* bernilai rendah, maka akan termasuk dalam kategori rendah. Jika hasil *XGBoost* bernilai menengah ke bawah, maka akan termasuk dalam kategori menengah ke bawah. Jika hasil *XGBoost* bernilai menengah ke atas, maka akan termasuk dalam kategori menengah ke atas. Terakhir, jika hasil *XGBoost* bernilai tinggi, maka akan termasuk dalam kategori tinggi. Kemudian menentukan prediksi terhadap model klasifikasi berdasarkan pohon klasifikasi, model klasifikasi yang telah terbentuk akan menghasilkan klasifikasi indeks ketimpangan gender pada 514 kabupaten/kota dalam tiga kelas. Jika hasil *XGBoost* bernilai rendah, maka akan termasuk dalam kategori rendah. Kemudian jika hasil *XGBoost* bernilai sedang, maka masuk dalam kategori sedang. Selanjutnya, apabila hasil *XGBoost* bernilai tinggi, maka akan termasuk dalam kategori tinggi. Langkah

selanjutnya, melakukan pengukuran akurasi menggunakan *confusion matrix* dan APER untuk melihat tingkat akurasi yang dihasilkan dari analisis klasifikasi.

2.8 Kajian Integrasi Ketimpangan Gender degan Al-Quran

Pemahaman terhadap masalah sosial dalam prespektif islam termasuk ketimpangan gender memiliki dampak signifikan dalam menganalisis siklus masyarakat di suatu wilayah dengan mayoritas penduduk beragama islam. Indonesia sebagai negara dengan mayoritas penduduk beragama islam ketimpangan gender menjadi permasalahan yang penting. Ketimpangan gender memengaruhi berbagai aspek masyaraat, sepeprti kesehatan, pendidikan, dan partisipasi politik. Upaya untuk mengurangi ketimpangan gender dalam masyarakat diharapkan tercapainya pembangunan manusia yang berkelanjutan serta menciptakan masyarakat yang lebih sehat, berpendidikan dan inklusif (Taufiq, 2007).

Pada Qur'an Kemenag (2022), ayat yang menunjukkan prinsip-prinsip kesetaraan antara laki-laki dan perempuan dalam Islam terdapat dalam Surah An-Nisa (4:32):

"Dan janganlah kamu iri hati terhadap apa yang telah Allah anugerahkan kepada sebagian kamu melebihi sebagian yang lain. Bagi laki-laki ada bahagian dari hasil usahanya, dan bagi perempuan (pun) ada bahagian dari hasil usahanya. Dan mohonlah kepada Allah sebagian dari karunia-Nya. Sesungguhnya Allah Maha Mengetahui segala sesuatu."

Ayat ini menjelaskan bahwa laki-laki dan perempuan mendapatkan bagian dan hak yang setara dalam hal tanggungjawab dan pemberian Allah. Ayat ini mengingatkan agar manusia tidak iri hati terhadap pemberian Allah kepada orang lain, karena setiap individu baik laki-laki maupun perempuan akan mendapatkan bagian dari hasil usahanya sendiri. Ini mencerminkan prinsip kesetaraan gender dalam Islam,

di mana baik laki-laki maupun perempuan memiliki tanggungjawab dan hak yang sama di hadapan Allah.

Imam Ahmad meriwayatkan dari Mujahid, ia berkata, Ummu Salamah berkata (Ghoffar dkk., 2001):

“Wahai Rasulullah! Kaum laki-laki dapat ikut serta berperang, sedangkan kami tidak diikutsertakan berperang dan hanya mendapat setengah bagian warisan. Dan janganlah kamu iri hati terhadap apa yang dikaruniakan Allah kepada sebahagianmu lebih banyak dari sebahagian yang lain.”

Surah An-Nisa ayat 32, dalam tafsir Al-Misbah karya M. Quraish Shihab, melarang perasaan iri hati terhadap anugerah yang Allah berikan kepada sesama, baik laki-laki maupun perempuan. Meskipun memiliki rezeki dan kelebihan yang berbeda, setiap manusia diciptakan dengan hak dan kewajiban dasar yang sama. Hal ini menegaskan bahwa setiap manusia memiliki potensi dan nilai yang sama, terlepas dari perbedaan jenis kelamin. Pada ayat ini menekankan kesetaraan hak dan tanggung jawab yang sama dihadapan Allah, hal ini menunjukkan bahwa laki-laki maupun perempuan memiliki hak yang sama untuk mendapatkan kesempatan dan perlakuan yang adil (Shihab, 2006).

Dalam perspektif islam, dijelaskan bahwa amal perbuatan yang dilakukan oleh laki-laki dan perempuan adalah setara. Setiap individu, baik laki-laki maupun perempuan berpotensi untuk memperoleh pahala dan dosa. Allah telah menetapkan hak-hak waris bagi keduanya. Amal kebaikan yang dilakukan oleh perempuan akan diberi pahala sepuluh kali lipat sama dengan pahala yang diperoleh seorang laki-laki. Dengan dasar ini, Allah memberikan prioritas kepada sebagian individu atas yang lain berdasarkan kesiapan dan tingkatan masing-masing individu. Perbedaan-

perbedaan keutamaan ini meliputi perbedaan lahiriah seperti karakter (akhlak), dan juga batiniah seperti ilmu dan kedudukan (Az Zuhaili, 2018).

Dapat disimpulkan bahwa islam menegaskan kesetaraan hak antar laki-laki dan perempuan dalam hal memperoleh pendidikan dan pengetahuan, hak untuk bekerja dan berpartisipasi dalam kehidupan sosial dan politik, dan hak untuk mendapatkan perlindungan hukum. Konsep kesetaraan gender dalam Islam yang menekankan pentingnya mendapatkan kesempatan yang sama untuk mengembangkan potensi dan kemampuan mereka. Oleh sebab itu, penting untuk meningkatkan pendidikan dan pemahaman tentang konsep kesetaraan gender dalam masyarakat. Hal ini dapat mengurangi ketimpangan gender dan dapat meningkatkan kesadaran dan pemahaman tentang hak-hak laki-laki dan perempuan.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dan studi literatur. Penelitian ini menggunakan data kuantitatif yaitu data Indeks Ketimpangan Gender pada 514 Kabupaten/Kota di Indonesia pada tahun 2022, kemudian dianalisis menggunakan metode *XGBoost*.

3.2 Data dan Sumber Data

Penelitian ini menggunakan data sekunder, yaitu data yang diperoleh dari sumber yang telah tersedia. Data tersebut diakses melalui situs resmi Badan Pusat statistika (BPS) Indonesia. Data yang digunakan meliputi variabel *dependent* (Y), yaitu data Indeks Ketimpangan Gender (IKG) tahun 2022 dan variabel *independent* (X). Berikut merupakan variabel yang digunakan dalam penelitian ini:

Tabel 3.1 Variabel Penelitian

Variabel		Skala	Keterangan
Y	Indeks Ketimpangan Gender (IKG) pada 514 kabupaten/Kota di Indonesia.	Ordinal	Rendah: ($IKG < 0,399$) Sedang: ($0,400 < IKG \leq 0,499$) Tinggi: ($IKG \geq 0,50$) Sumber: BPS, 2022
X_1	Persentase Persalinan Tidak Difasilitasi Kesehatan	Rasio	0 – 100
X_2	Persentase Wanita Berusia 15-49 Tahun Yang Pernah Kawin dan Saat Melahirkan Hidup Pertama Berusia Kurang Dari 20 Tahun	Rasio	0 – 100

Variabel		Skala	Keterangan
X_3	Persentase Laki-Laki yang Duduk di Parlemen	Rasio	0 – 100
X_4	Persentase Perempuan yang Duduk di Parlemen	Rasio	0 – 100
X_5	Persentase Penduduk Laki-laki dengan Pendidikan Minimal SMA	Rasio	0 – 100
X_6	Persentase Penduduk Perempuan dengan Pendidikan minimal SMA	Rasio	0 – 100
X_7	Persentase Angkatan Kerja Laki-laki Terhadap Penduduk Usia Kerja	Rasio	0 – 100
X_8	Persentase Angkatan Kerja Perempuan Terhadap Penduduk Usia Kerja	Rasio	0 – 100

3.3 Tahapan Penelitian

Penelitian ini menggunakan metode *XGBoost* dengan tahapan penelitian sebagai berikut:

1. Mengumpulkan data klasifikasi IKG tahun 2022, yang terdiri dari variabel *dependent* dan variabel *independent*.
2. Melakukan analisis deskriptif terhadap data IKG untuk mendapatkan gambaran pada data IKG.
3. Membagi data kedalam data *training* dan data *testing*. Data *training* digunakan untuk pemodelan dan data *testing* digunakan untuk evaluasi performa model. Pembagian data *testing* dan data *training* berdasarkan percobaan *splitting* data sebanyak lima kali yaitu (70%:30%), (75%:25%), (80%:20%), (85%:15%) dan (90%:10%) dengan tingkat akurasi terbaik yaitu kombinasi persentase 90% data *training* dan 10% data *testing*.
4. Pembentukan pohon klasifikasi menggunakan metode CART
 - a. Pemilihan variabel pemilah pada variabel independen dilakukan menggunakan kaidah *goodness of split* dan indeks gini. Variabel yang

memiliki nilai *goodness of split* tertinggi akan ditetapkan sebagai *parent node*.

Persamaan indeks gini adalah sebagai berikut:

$$i(t) = 1 - \sum_{m=1}^k p^2(m|t)$$

Persamaan *goodness of split* sebagai berikut:

$$\begin{aligned} \theta(s, t) &= \Delta i(s, t) \\ &= i(t) - P_L \cdot i(t_L) - P_R \cdot i(t_R) \end{aligned}$$

- b. Penentuan terminal *node* (simpul akhir)
- c. Pelabelan kelas pada terminal *node*, pelabelan ini mengacu pada nilai mengacu pada nilai $N(m|t)$. Nantinya label kelas berdasarkan nilai $P(m|t)$ yang paling besar. berikut merupakan persamaan dalam menentukan kelas:

$$p(m_0|t) = \max p(m|t) = \max \frac{N_m(t)}{N(t)}$$

- d. Pemangkasan pohon klasifikasi
 - e. Penentuan pohon klasifikasi optimal
5. Tahapan pemodelan menggunakan metode *XGBoost*, menentukan *hyperparameter* guna menentukan nilai parameter optimal dan menghindari terjadinya *overfitting*.
 6. Melakukan evaluasi model pada data *testing* menggunakan *confusion matrix* dan APER dengan persamaan sebagai berikut:

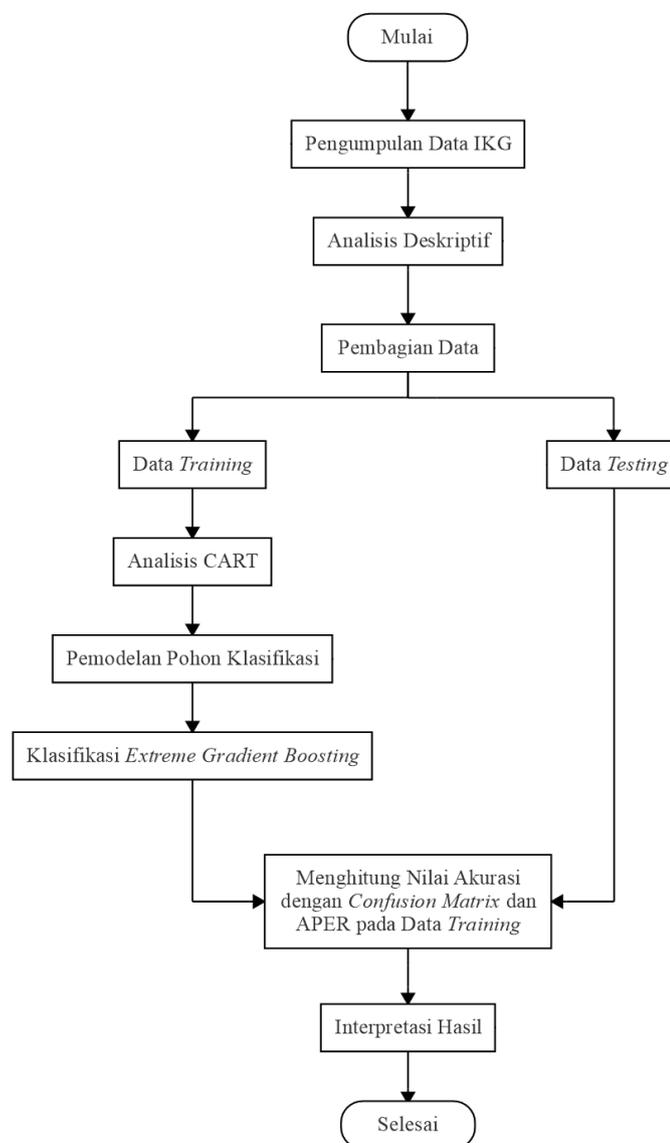
$$APER = \left(\frac{FP + FN}{TP + TN + FP + FN} \right) \times 100$$

persamaan tingkat akurasi sebagai berikut:

$$\text{Tingkat akurasi} = 1 - \text{APER}$$

3.4 Flowchart Penelitian

Berikut adalah representasi grafis dari tahapan penelitian menggunakan metode *XGBoost* dalam penentuan klasifikasi Indeks Ketimpangan Gender di Indonesia.



Gambar 3.1 Flowchart Penelitian

BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisis Statistik Deskriptif

Data penelitian mengenai Indeks Ketimpangan Gender (IKG) pada 514 Kabupaten/Kota di Indonesia tahun 2022 menunjukkan bahwa semakin rendah nilai IKG suatu Kabupaten/Kota, maka kesetaraan gender di wilayah tersebut akan semakin baik. Suatu Kabupaten/Kota memiliki kategori IKG kelas rendah jika memiliki IKG pada rasio kurang dari 0,399. Apabila IKG berada pada rasio 0,400 – 0,449, maka memiliki kategori IKG sedang. Kemudian suatu Kabupaten/Kota memiliki kategori kelas tinggi jika memiliki IKG pada rasio lebih dari 0,50.



Gambar 4.1 *Pie Chart* IKG 2022

Gambar 4.1 menunjukkan *pie chart* data IKG pada Kabupaten/Kota di Indonesia pada tahun 2022. Pada *pie chart* tersebut yang memiliki rentang kelas rendah sebanyak 148 Kabupaten/Kota atau sebesar 29%, kelas sedang sebanyak 136 Kabupaten/Kota atau sebesar 26%, dan kelas tinggi sebanyak 230

Kabupaten/Kota atau sebesar 45%. Terdapat beberapa indikator yang memengaruhi kelas pada nilai IKG menurut Kabupaten/Kota di Indonesia tahun 2022 di Indonesia. Indikator-indikator tersebut menjadi variabel *independent* (X) dalam penelitian ini, diantaranya adalah indikator persalinan tidak difasilitasi kesehatan (X_1), persentase wanita berusia 15-49 tahun yang pernah kawin dan saat melahirkan hidup pertama berusia kurang dari 20 tahun (X_2), persentase laki-laki yang duduk di parlemen (X_3), perempuan yang duduk di parlemen (X_4), penduduk laki-laki dengan pendidikan minimal SMA (X_5), penduduk perempuan dengan pendidikan minimal SMA (X_6), angkatan kerja perempuan terhadap penduduk usia kerja (X_7), angkatan kerja perempuan terhadap penduduk usia kerja (X_8). Indikator pertama yang memengaruhi nilai indeks ketimpangan gender di Indonesia yaitu persalinan tidak difasilitasi kesehatan (X_1). Berikut analisis deskriptif berdasarkan kategori kelas pada indikator pertama yaitu `persalinan tidak difasilitasi kesehatan (X_1).

Tabel 4.1 Analisis Deskriptif Persentase Persalinan Tidak Difasilitasi Kesehatan (X_1)

Kelas	Mean	Std. Deviasi	Minimum	Maksimum
Tinggi	32,40	18,40	0,00	100
Sedang	12,29	7,24	2,00	100
Rendah	2,70	2,62	0,00	100

Berdasarkan tabel 4.1 indikator persentase persalinan tidak difasilitasi kesehatan (X_1) menunjukkan persentase perempuan di kabupaten/kota yang melahirkan tanpa bantuan fasilitas kesehatan. Pada kategori tinggi diperoleh rata-rata X_1 sebesar 32,40% persalinan di kabupaten/kota tidak difasilitasi kesehatan, artinya sekitar 32 dari setiap 100 perempuan melahirkan tanpa difasilitasi kesehatan. Standar deviasi sebesar 18,40 menandakan variasi yang cukup besar

antar daerah pada kategori tinggi. Nilai minimum dan maksimum pada ketiga kategori berkisar antara 0 sampai 100, menunjukkan adanya ketimpangan antar daerah. Dimana terdapat daerah yang sepenuhnya persalinan difasilitasi kesehatan, namun masih ada daerah yang sepenuhnya persalinan tidak difasilitasi kesehatan. Dari analisis deskriptif indikator persalinan tidak difasilitasi kesehatan (X_1) dapat disimpulkan bahwa adanya ketimpangan terhadap akses fasilitas kesehatan, khususnya dalam proses persalinan. Kabupaten/kota yang memiliki kategori IKG tinggi cenderung memiliki akses yang lebih rendah terhadap persalinan difasilitasi kesehatan.

Indikator kedua yaitu persentase wanita berusia 15-49 tahun yang pernah kawin dan saat melahirkan hidup pertama berusia kurang dari 20 tahun (X_2). Berikut analisis deskriptif berdasarkan kategori kelas pada indikator (X_2).

Tabel 4.2 Analisis Deskriptif Persentase Wanita Berusia 15-49 Tahun yang Pernah Kawin dan Saat Melahirkan Hidup Pertama Berusia Kurang dari 20 Tahun (X_2)

Kelas	Mean	Std. Deviasi	Minimum	Maksimum
Tinggi	31,93	8,30	0,20	52,80
Sedang	27,47	8,22	0,20	50,40
Rendah	19,34	8,24	4,30	44,90

Berdasarkan tabel 4.2 menggambarkan analisis deskriptif indikator kedua, yaitu persentase wanita berusia 15-49 tahun yang pernah kawin dan saat melahirkan hidup pertama berusia kurang dari 20 tahun (X_2). Pada kategori tinggi diperoleh rata-rata X_2 sebesar 31,93% perempuan di kabupaten/kota pernah melahirkan pertama kali kurang dari 20 tahun. Standar deviasi sebesar 8,30 menunjukkan adanya variasi atau perbedaan yang cukup signifikan antar daerah pada kategori tinggi. Pada kategori tinggi dan sedang tercatat nilai minimum sebesar 0,20, yang

menandakan bahwa terdapat daerah yang hampir tidak memiliki kasus kelahiran pada usia muda. Sementara itu, nilai maksimum mencapai 52,80 pada kategori tinggi, yang berarti lebih dari setengah perempuan di suatu daerah mengalami kelahiran pertama di usia remaja. Berdasarkan analisis deskriptif indikator persentase wanita berusia 15-49 tahun yang pernah kawin dan saat melahirkan hidup pertama berusia kurang dari 20 tahun (X_2) dapat disimpulkan bahwa kabupaten/kota kategori tinggi cenderung memiliki nilai rata-rata lebih tinggi dibandingkan dengan kategori sedang dan rendah. Hal ini menunjukkan bahwa kabupaten/kota dengan IKG tinggi terlihat lebih banyak perempuan yang melahirkan kurang dari 20 tahun. Hal tersebut menunjukkan adanya ketimpangan gender terhadap akses pendidikan dan kesehatan reproduksi.

Selanjutnya indikator persentase laki-laki yang duduk di parlemen (X_3). Berikut analisis deskriptif berdasarkan kategori kelas pada indikator persentase laki-laki yang duduk di parlemen (X_3)

Tabel 4.3 Analisis Deskriptif Persentase Laki-laki yang Duduk di Parlemen (X_3)

Kelas	Mean	Std. Deviasi	Minimum	Maksimum
Tinggi	88,56	8,35	52,00	100
Sedang	81,95	7,86	52,94	96,67
Rendah	80,63	8,03	51,43	96,67

Berdasarkan tabel 4.3 menggambarkan analisis deskriptif indikator persentase laki-laki yang duduk di parlemen (X_3). Pada kategori tinggi diperoleh rata-rata sebesar 88,56%, artinya anggota parlemen di kabupaten/kota dengan kategori tinggi diduduki oleh 89 laki-laki dari setiap 100 penduduk. Standar deviasi sebesar 8,35 menunjukkan adanya variasi antar kabupaten/kota pada kategori tinggi. Nilai

minimum sebesar 52,00 dan maksimum sebesar 100 pada kategori tinggi mengindikasikan bahwa di seluruh daerah tersebut, laki-laki menjadi mayoritas anggota parlemen, bahkan terdapat daerah dengan anggota parlemen seluruhnya dari laki-laki. Kategori sedang memiliki rata-rata sebesar 81,95 dengan standar deviasi 7,86, sedangkan kategori rendah memiliki rata-rata 80,63 dengan standar deviasi 8,03. Nilai minimum pada kategori sedang dan rendah di atas 50, yang menandakan bahwa laki-laki secara konsisten lebih dominan dibanding perempuan dalam struktur parlemen. Hal ini mencerminkan adanya ketimpangan gender dalam partisipasi politik di tingkat legislatif.

Selanjutnya indikator persentase perempuan yang duduk di parlemen (X_4). Berikut analisis deskriptif berdasarkan kategori kelas pada indikator persentase persentase perempuan yang duduk di parlemen (X_4).

Tabel 4.4 Analisis Deskriptif Persentase Perempuan yang Duduk di Parlemen (X_4)

Kelas	Mean	Std. Deviasi	Minimum	Maksimum
Tinggi	11,44	8,35	0,00	48,00
Sedang	18,05	7,86	3,33	47,06
Rendah	19,37	8,03	3,33	48,57

Berdasarkan tabel 4.4 menggambarkan analisis deskriptif indikator persentase perempuan yang duduk di parlemen (X_4). Pada kategori tinggi diperoleh rata-rata sebesar 11,44% perempuan yang duduk di parlemen, artinya dari setiap 100 anggota parlemen pada kategori tinggi, hanya sekitar 11 orang yang merupakan perempuan. Standar deviasi sebesar 8,35 menunjukkan adanya perbedaan yang cukup besar antar daerah dalam hal keterwakilan perempuan yang duduk di parlemen. Nilai minimum sebesar 0,00 menandakan bahwa terdapat daerah yang sama sekali tidak

memiliki perempuan sebagai anggota parlemen, sementara nilai maksimum sebesar 48,00 menunjukkan daerah dengan keterwakilan perempuan hampir setengah dari total anggota parlemen. Pada kategori sedang, rata-rata persentase perempuan di parlemen meningkat menjadi 18,05 dan pada kategori rendah sebesar 19,37, dengan standar deviasi yang serupa. Meskipun secara umum keterwakilan perempuan lebih tinggi di kategori sedang dan rendah, keseluruhan data tetap menunjukkan bahwa representasi perempuan masih jauh lebih rendah dibandingkan laki-laki yang duduk di parlemen. Hal ini mencerminkan ketimpangan gender yang cukup signifikan dalam partisipasi politik dan pengambilan keputusan di tingkat legislatif daerah.

Selanjutnya indikator persentase penduduk laki-laki dengan pendidikan minimal SMA (X_5). Berikut analisis deskriptif berdasarkan kategori kelas pada indikator persentase penduduk laki-laki dengan pendidikan minimal SMA (X_5).

Tabel 4.5 Analisis Deskriptif Persentase Penduduk Laki-laki dengan Pendidikan Minimal SMA (X_5)

Kelas	Mean	Std. Deviasi	Minimum	Maksimum
Tinggi	35,88	9,61	13,45	65,17
Sedang	39,68	12,52	20,07	77,28
Rendah	50,77	16,10	8,53	83,85

Berdasarkan tabel 4.5 menggambarkan analisis deskriptif indikator persentase persentase penduduk laki-laki dengan pendidikan minimal SMA (X_5). Pada kategori tinggi diperoleh rata-rata sebesar 35,88%, artinya dari 100 laki-laki sekitar 36 orang telah menyelesaikan pendidikan setidaknya sampai tingkat SMA. Standar deviasi pada kategori tinggi sebesar 9,61 menunjukkan adanya variasi antar daerah. Nilai minimum pada kategori tinggi tercatat sebesar 13,45 dan maksimum sebesar 65,17. Pada kategori sedang, rata-rata meningkat menjadi 39,68 dengan

nilai minimum 20,07 dan maksimum 77,28. Sementara itu, pada kategori rendah rata-rata sebesar 50,77 dengan nilai minimum 8,53 dan maksimum 83,85. Hal ini menunjukkan bahwa daerah dengan IKG rendah memiliki rata-rata pendidikan laki-laki dengan pendidikan minimal pendidikan SMA yang lebih tinggi.

Tabel 4.6 Analisis Deskriptif Persentase Penduduk Perempuan dengan Pendidikan Minimal SMA (X_6)

Kelas	Mean	Std. Deviasi	Minimum	Maksimum
Tinggi	35,69	9,98	4,51	62,11
Sedang	35,64	12,10	15,69	71,98
Rendah	46,57	16,13	3,65	78,21

Berdasarkan Tabel 4.6 menggambarkan analisis deskriptif indikator persentase penduduk perempuan dengan pendidikan minimal SMA (X_6). Pada kategori tinggi diperoleh nilai rata-rata sebesar 35,69, artinya dari setiap 100 perempuan hanya 36 orang telah menempuh pendidikan minimal tingkat SMA. Standar deviasi sebesar 9,98 menunjukkan adanya variasi yang cukup besar antar daerah pada ategori tinggi dengan nilai minimum 4,51 dan maksimum 62,11. Berdasarkan analisis deskriptif indikator persentase penduduk perempuan dengan pendidikan minimal SMA (X_6) dapat disimpulkan bahwa kabupaten/kota kategori rendah memiliki nilai rata-rata persentase perempuan lebih tinggi dibandingkan dengan kategori sedang dan tinggi. Hal ini menunjukkan bahwa kabupaten/kota dengan IKG rendah terlihat lebih banyak perempuan dengan pendidikan minimal SMA.

Selanjutnya indikator persentase angkatan kerja laki-laki terhadap penduduk usia kerja (X_7). Berikut analisis deskriptif berdasarkan kategori kelas

pada indikator persentase angkatan kerja laki-laki terhadap penduduk usia kerja (X_7).

Tabel 4.7 Analisis Deskriptif Persentase Angkatan Kerja Laki-laki Terhadap Penduduk Usia Kerja (X_7)

Kelas	Mean	Std. Deviasi	Minimum	Maksimum
Tinggi	84,99	3,78	72,41	97,72
Sedang	84,30	3,59	74,36	90,74
Rendah	82,43	4,32	69,37	98,55

Berdasarkan Tabel 4.7 menggambarkan analisis deskriptif indikator persentase angkatan kerja laki-laki terhadap penduduk usia kerja (X_7). Pada kategori tinggi diperoleh nilai rata-rata sebesar sebesar 84,99, artinya sekitar 85 dari setiap 100 laki-laki usia kerja di daerah tersebut tergolong angkatan kerja. Standar deviasi sebesar 3,78 menandakan bahwa perbedaan antar daerah dalam kategori tinggi ini tidak terlalu besar. Nilai minimum adalah 72,41 dan maksimum 97,72, yang berarti hampir seluruh laki-laki usia kerja terlibat dalam pasar kerja, namun juga ada daerah yang partisipasinya lebih rendah. Secara keseluruhan tingkat partisipasi kerja laki-laki cukup tinggi di seluruh kategori IKG, meskipun masih adanya variasi antar daerah. Hal ini dapat menunjukkan bahwa laki-laki memiliki akses dan keterlibatan dalam kegiatan ekonomi.

Selanjutnya indikator persentase angkatan kerja perempuan terhadap penduduk usia kerja (X_8). Berikut analisis deskriptif berdasarkan kategori kelas pada indikator persentase angkatan kerja perempuan terhadap penduduk usia kerja (X_8).

Tabel 4.8 Analisis Deskriptif Persentase Angkatan Kerja Perempuan Terhadap Penduduk usia Kerja (X_8)

Kelas	Mean	Std. Deviasi	Minimum	Maksimum
Tinggi	55,62	13,30	29,72	98,28
Sedang	54,86	9,06	36,31	81,33
Rendah	56,29	10,06	34,88	96,26

Berdasarkan Tabel 4.8 menggambarkan analisis deskriptif indikator persentase angkatan kerja perempuan terhadap penduduk usia kerja (X_8). Pada kategori tinggi diperoleh rata-rata sebesar 55,62 yang berarti sekitar 56 dari setiap 100 perempuan usia kerja aktif dalam angkatan kerja. Nilai standar deviasi sebesar 13,30 menunjukkan variasi yang cukup besar antar daerah pada kategori tinggi. Pada kategori tinggi nilai minimum sebesar 29,72 mengindikasikan adanya daerah di mana kurang dari sepertiga perempuan usia kerja terlibat dalam pasar kerja, sementara nilai maksimum sebesar 98,28 menunjukkan bahwa di beberapa daerah hampir seluruh perempuan usia kerja masuk ke dalam angkatan kerja. Pada kategori IKG sedang, rata-rata partisipasi perempuan adalah 54,86 dengan standar deviasi 9,06. Sementara itu, pada kategori IKG rendah, rata-rata partisipasi perempuan sedikit lebih tinggi yaitu 56,29 dengan variasi antara 34,88. Berdasarkan analisis deskriptif indikator persentase angkatan kerja perempuan terhadap penduduk usia kerja (X_8) dapat disimpulkan bahwa tingkat partisipasi perempuan dalam angkatan kerja masih lebih rendah dibandingkan laki-laki dan terdapat kesenjangan antar daerah yang cukup besar, terutama pada kategori IKG tinggi. Hal ini mencerminkan adanya ketimpangan dalam hal akses perempuan terhadap lapangan pekerjaan.

4.2 Analisis CART

Analisis CART digunakan untuk mengetahui hasil klasifikasi pada data indeks ketimpangan gender menurut kabupaten/kota di Indonesia tahun 2022 yang dikelompokkan dalam tiga kategori yaitu rendah, sedang, dan tinggi. Langkah awal pada penelitian ini adalah membagi data menjadi dua bagian, yaitu data *training* dan data *testing* dengan berbagai kombinasi persentase terhadap 514 data IKG tahun 2022.

4.2.1 Penentuan Data *Training* dan Data *Testing*

Pada penelitian ini, kombinasi pembagian data training dan data testing pada penelitian ini dibagi menjadi 4 kombinasi, yaitu persentase (75%:25%), (80%:20%), (85%:15%), dan (90%:10). Selanjutnya, model dibangun berdasarkan setiap kombinasi persentase data dan tingkat akurasi klasifikasinya dihitung. Hasil akurasi dari masing-masing kombinasi dibandingkan untuk menentukan kombinasi yang menghasilkan akurasi paling seimbang antara data training dan data testing. Data training akan digunakan untuk membangun model pohon klasifikasi, sedangkan data testing akan digunakan untuk mengukur performa model.

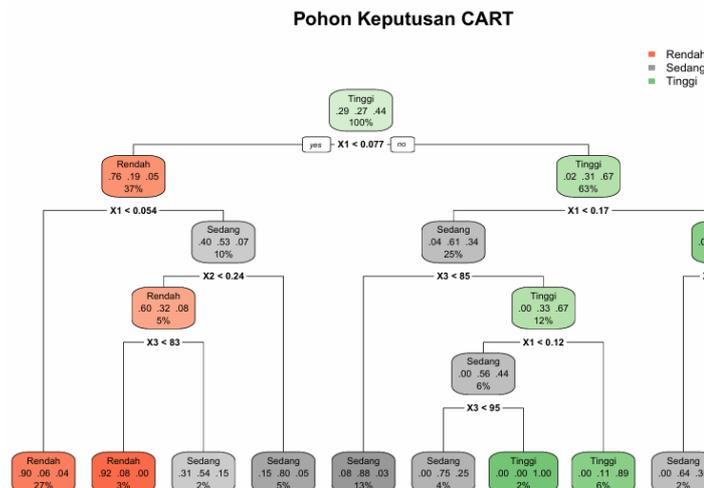
Tabel 4.9 Pembagian Data *Training* dan Data *Testing*

Persentase data (%)		Akurasi klasifikasi	
<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
75	25	0,914	0,871
80	20	0,92	0,883
85	15	0,925	0,892
90	10	0,93	0,905

Tabel 4.9 menunjukkan hasil akurasi klasifikasi berdasarkan persentase data *training* dan data *testing*. Kombinasi yang menghasilkan seimbang antara data *training* dan data *testing* ditemukan pada kombinasi persentase 90% untuk data *training* dan 10% untuk data *testing*, kombinasi ini memiliki selisih akurasi sebesar 0,025. Oleh karena itu, penelitian selanjutnya menggunakan kombinasi persentase (90%:10%) dengan jumlah data *training* sebanyak 463 dan data *testing* berjumlah 51.

4.2.2 Pembentukan Pohon Klasifikasi CART

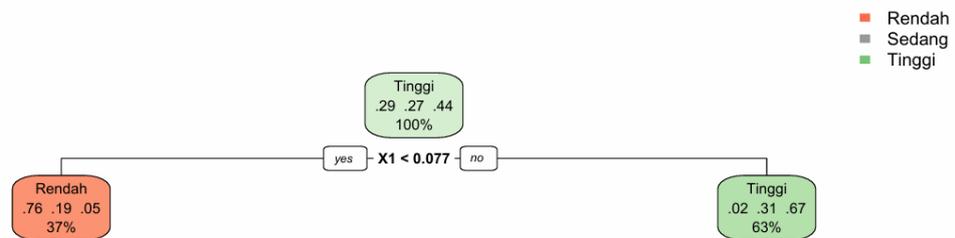
Berdasarkan hasil pemodelan menggunakan metode CART (*Classification and Regression Tree*), diperoleh struktur pohon klasifikasi data IKG 514 Kabupaten/kota di Indonesia tahun 2022 dalam tiga kategori kelas, yaitu rendah, sedang, dan tinggi. Struktur lengkap dari pohon klasifikasi CART ditunjukkan pada gambar berikut.



Gambar 4.2 Hasil Pohon Klasifikasi pada IKG Kabupaten/Kota Tahun 2022

Gambar 4.2 merupakan hasil pohon klasifikasi berdasarkan data IKG kabupaten/kota di Indonesia pada tahun 2022 dengan *root node* yang memiliki label tinggi dikarenakan sebanyak 44% data *training* menunjukkan IKG yang memiliki kategori tinggi. Kemudian dilakukan pembagian simpul menjadi simpul kanan dan simpul kiri. Pada simpul kanan dihasilkan 64% data dengan Indikator persalinan tidak difasilitasi kesehatan (X_1), lebih kecil dari 0,077, sedangkan pada simpul kiri dihasilkan 36% data dengan indikator persalinan tidak difasilitasi kesehatan (X_1), lebih besar sama dengan 0,077. Kedua simpul tersebut kemudian masih dibagi lagi menjadi internal node selanjutnya sampai didapatkan leaf node yang merupakan simpul keputusan akhir dan sudah homogen. Dengan pohon klasifikasi yang telah dihasilkan, kita dapat melakukan perhitungan manual dengan rincian penghitungan indeks keheterogenan simpul awal menggunakan rumus indeks gini, pengukuran *goodness of split* pada setiap percabangan, serta

penentuan label kelas pada setiap simpul daun berdasarkan probabilitas tertinggi dari masing-masing kategori.



Gambar 4.3 Pohon Klasifikasi CART

Pembentukan pohon klasifikasi diawali dengan menghitung nilai keheterogenan simpul awal menggunakan fungsi keheterogenan indeksgini. Kemudian untuk memilah simpul awal menjadi dua simpul anak, maka dilakukan penentuan pemilah dari semua kemungkinan nilai variabel (*threshold*). Berikut perhitungan manual indeks gini:

$$p(\text{rendah}) = p(1|t) = \frac{148}{514} = 0,29$$

$$p(\text{sedang}) = p(2|t) = \frac{139}{514} = 0,27$$

$$p(\text{tinggi}) = p(3|t) = \frac{227}{514} = 0,44$$

Secara umum persamaan indeks gini yang telah didefinisikan pada persamaan (2.1) sebagai berikut:

$$i(t) = \sum_{n=1, n \neq m}^k p(n|t) p(m|t)$$

$$p(1|t) \cdot p(2|t) = 0,29 \cdot 0,27 = 0,078$$

$$p(1|t) \cdot p(3|t) = 0,29 \cdot 0,44 = 0,127$$

$$p(2|t) \cdot p(1|t) = 0,27 \cdot 0,29 = 0,078$$

$$p(2|t) \cdot p(3|t) = 0,27 \cdot 0,44 = 0,118$$

$$p(3|t) \cdot p(1|t) = 0,44 \cdot 0,29 = 0,127$$

$$p(3|t) \cdot p(2|t) = 0,44 \cdot 0,27 = 0,118$$

$$i(t) = 0,078 + 0,127 + 0,078 + 0,118 + 0,127 + 0,118$$

$$i(t) = 0,649$$

Nilai indeks gini berkisar antara 0 sampai 1, dimana apabila nilainya mendekati 0 artinya pemisahan semakin sempurna. Berdasarkan perhitungan di atas dihasilkan nilai indeks gini 0,649, sehingga perlu dilakukan pemisahan berikutnya guna dihasilkan pemisahan yang semakin baik. Kemudian untuk memilah simpul awal menjadi dua simpul anak, maka dilakukan penentuan pemilah dari nilai variabel (*threshold*). Terlebih dahulu menghitung nilai *goodness of split*, sebagai berikut.

Probabilitas untuk masing-masing kategori pada *node* kiri:

Rendah: 76% = 144 data

Sedang: 19% = 36 data

Tinggi: 5% = 10 data

$$\begin{aligned} i(t_L) &= 1 - \sum_{m=1}^k p^2(n|t) \\ &= 1 - \left(\left(\frac{144}{190} \right)^2 + \left(\frac{36}{190} \right)^2 + \left(\frac{10}{190} \right)^2 \right) \\ &= 1 - \left((0,758)^2 + (0,189)^2 + (0,053)^2 \right) \end{aligned}$$

$$\begin{aligned}
&= 1-(0,574+0,036+0,003) \\
&= 1-0,613 \\
&= 0,387
\end{aligned}$$

Probabilitas untuk masing-masing kategori pada *node* kanan:

Rendah: 2% = 7 data

Sedang: 31% = 100 data

Tinggi: 67% = 217 data

$$\begin{aligned}
i(t_R) &= 1-\sum_{m=1}^k P^2(m|t) \\
&= 1-\left(\left(\frac{7}{324}\right)^2 + \left(\frac{100}{324}\right)^2 + \left(\frac{217}{324}\right)^2\right) \\
&= 1-\left((0,022)^2 + (0,309)^2 + (0,670)^2\right) \\
&= 1-(0,001+0,095+0,449) \\
&= 1-0,545 \\
&= 0,455
\end{aligned}$$

Selanjutnya menentukan peluang objek pada *node* kiri dan *node* kanan menggunakan persamaan (2.4) dan (2.5), sebagai berikut:

$$\begin{aligned}
P_L &= \frac{N(t_L)}{N} & \text{dan} & & P_R &= \frac{N(t_R)}{N} \\
&= \frac{190}{514} & & & &= \frac{324}{514} \\
&= 0,369 & & & &= 0,630
\end{aligned}$$

Sehingga didapatkan nilai *goodness of split* pada persamaan (2.3), sebagai berikut:

$$\begin{aligned}
\phi(s,t) &= \Delta i(s,t) \\
&= i(t) - P_L \cdot i(t_L) - P_R \cdot i(t_R) \\
&= 0,649 - (0,369 \cdot 0,387) - (0,630 \cdot 0,455) \\
&= 0,649 - 0,143 - 0,287 \\
&= 0,219
\end{aligned}$$

Selanjutnya penandaan label kelas menggunakan persamaan (2.6):

$$\begin{array}{ll}
p(1|t_L) = \frac{129}{182} = 0,709 & \text{dan} & p(1|t_R) = \frac{5}{291} = 0,017 \\
p(2|t_L) = \frac{35}{182} = 0,192 & & p(2|t_R) = \frac{90}{291} = 0,309 \\
p(3|t_L) = \frac{8}{182} = 0,044 & & p(3|t_R) = \frac{196}{291} = 0,674
\end{array}$$

Sehingga diketahui bahwa label kelas untuk node kiri adalah kelas 1 (rendah) karena $p(1|t_L) = 0,709$ adalah yang terbesar dan label kelas untuk node kanan adalah kelas 3 (tinggi) karena $p(1|t_R) = 0,674$ adalah yang terbesar. Selanjutnya pemangkasan pohon klasifikasi menggunakan menggunakan persamaan (2.7).

$$\begin{aligned}
R_\alpha(T) &= R(T) + \alpha [\tilde{T}] \\
&= 0,298 + (0,05 \cdot 8) \\
&= 0,298 + 0,4 \\
&= 0,698
\end{aligned}$$

Karena risiko sebelum pemangkasan $R_\alpha(T) = 0,698$ lebih besar daripada risiko sebelum pemangkasan $R(t) = 0,298$ pemangkasan tidak seharusnya dilakukan. Dalam hal ini, pemangkasan justru akan meningkatkan risiko atau kesalahan klasifikasi, yang bertentangan dengan tujuan pemangkasan untuk

mengurangi risiko dan meningkatkan generalisasi. Jadi, tidak dilakukan pemangkasan pada pohon klasifikasi ini karena risiko setelah pemangkasan lebih besar daripada risiko sebelum pemangkasan. Tahap terakhir yaitu penentuan pohon klasifikasi optimal menggunakan persamaan (2.8).

$$R^{ts}(T_t) = \frac{1}{N_2} \sum_{(x_n, m_n) \in I_2} X(d(x_n) \neq m_n)$$

$$R^{ts}(T_t) = \frac{1}{463} \cdot 259$$

$$R^{ts}(T_t) = 0,5594$$

Perhitungan menunjukkan bahwa risiko uji (*test risk*) adalah 0,5594 atau 55,94%, yang berarti bahwa 55,94% dari prediksi dalam data *training* salah klasifikasinya. Dengan risiko uji 55,94% pada model CART, diperlukan peningkatan performa model.

4.2.3 Tingkat Akurasi Ketepatan Klasifikasi CART

Tingkat akurasi dari pohon klasifikasi IKG yang dihasilkan digunakan untuk menetapkan kategori rendah, sedang, atau tinggi pada Kabupaten/Kota di Indonesia. Tingkat akurasi model dapat dievaluasi menggunakan nilai APER (*apparent error rate*) dan 1-APER (*total accuracy rate*). Tingkat ketepatan hasil klasifikasi IKG pada metode CART dapat diukur menggunakan data *testing* yang ditampilkan dalam tabel *confusion matrix* berikut.

Tabel 4.10 *Confusion Matrix* Klasifikasi CART

Pengamatan	Prediksi			Total
	Rendah	Sedang	Tinggi	
Rendah	13	1	0	14
Sedang	2	7	2	11
Tinggi	0	5	22	27
Total	15	13	24	52

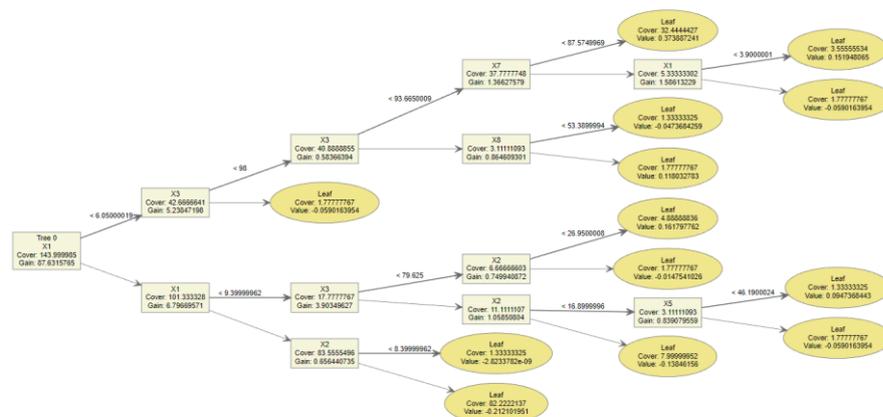
Tabel 4.10 menunjukkan hasil *confussion matrix* dari klasifikasi Indeks Ketimpangan Gender (IKG) menggunakan metode CART (*Classification and Regression Tree*) pada data *testing*. Berdasarkan tabel tersebut, terdapat 13 Kabupaten/Kota yang memiliki kategori IKG rendah berhasil diklasifikasikan dengan benar oleh model, sementara 3 Kabupaten/Kota dari kategori rendah salah diklasifikasikan ke kategori sedang. Pada kategori sedang, model mampu memprediksi dengan benar sebanyak 7 Kabupaten/Kota, tetapi terdapat 1 Kabupaten/Kota dari kategori sedang yang salah diklasifikasikan sebagai kategori rendah, serta 5 Kabupaten/Kota dari kategori sedang salah diklasifikasikan ke kategori tinggi. Untuk kategori tinggi, model menunjukkan performa yang sangat baik dengan 21 Kabupaten/Kota berhasil diklasifikasikan dengan benar, sementara hanya 1 Kabupaten/Kota yang salah diklasifikasikan ke kategori sedang. Kemudian perhitungan manual pada data *testing* adalah sebagai berikut:

$$\begin{aligned}
 APER &= \frac{FP + FN}{TP + TN + FP + FN} \cdot 100\% \\
 &= \frac{2 + 1 + 5 + 2}{52} \cdot 100\% \\
 &= 19,23\% \\
 akurasi &= 100 - APER \\
 &= 80,77\%
 \end{aligned}$$

Nilai akurasi dari hasil klasifikasi IKG menggunakan metode CART memiliki akurasi sebesar 80,39%, yang berarti dari 51 pengamatan sebanyak 41 pengamatan berhasil diklasifikasikan dengan benar oleh model. Pada model CART, diperlukan peningkatan performa model menggunakan algoritma yang lebih canggih seperti metode *XGBoost* untuk meningkatkan akurasi yang lebih baik. Penggunaan metode *XGBoost* diharapkan dapat meningkatkan akurasi model dan mengurangi risiko kesalahan klasifikasi dalam memprediksi data baru secara lebih akurat.

4.3 Analisis XGBoost

Pada penerapan metode *XGBoost*, algoritma ini menggunakan model klasifikasi melalui pendekatan *ensemble learning* dengan memanfaatkan *decision tree* sebagai *base learner*. Setiap *tree* umumnya memiliki struktur yang berbeda, hal ini dikarenakan setiap *tree* dilakukan secara bertahap dan iteratif, di mana setiap *tree* baru dibangun untuk memperbaiki kesalahan prediksi dari *tree* sebelumnya. Ilustrasi salah satu *tree* yang dihasilkan dalam model *XGBoost* diperlihatkan pada Gambar 4.4.



Gambar 4.4 Pohon Klasifikasi *XGBoost*

Pohon klasifikasi *XGBoost* pada gambar 4.4 menggambarkan metode *XGBoost* memodelkan pohon klasifikasi berbasis *decision tree* untuk memisahkan data IKG berdasarkan indikator yang paling berpengaruh dalam menentukan kategori rendah, sedang, dan tinggi. Struktur pohon ini terdiri dari *parent node* (simpul induk) dan *child node* (simpul anak) yang saling terhubung. Di setiap *node*, data dipisahkan berdasarkan kondisi nilai dari suatu variabel yang memiliki nilai *gain* tertinggi, yaitu kontribusi terbesar terhadap penurunan nilai *loss function*. Pada *root node*, model melakukan pemisahan data berdasarkan variabel indikator persalinan tidak difasilitasi kesehatan karena menunjukkan X_1 memiliki kontribusi terbesar terhadap peningkatan akurasi model dalam memisahkan kelas yang ditunjukkan oleh nilai *gain* tinggi sebesar 87,63.

Kondisi pencabangan pada *root node* adalah $X_1 < 6,05$ yang membagi data menjadi dua cabang utama, yaitu cabang kiri untuk data yang memenuhi $X_1 < 6,05$ dan cabang kanan untuk data dengan $X_1 \geq 6,05$. Setelah pembelahan di *parent node*, terbentuklah *child node* yang melakukan pencabangan lebih lanjut berdasarkan variabel berikutnya dengan nilai *gain* tertinggi. Proses ini berlanjut hingga data mencapai simpul akhir (*leaf*). *Leaf node* pada pohon ini berisi nilai *value* yang merepresentasikan kelas prediksi akhir. Sementara itu, nilai *cover* menunjukkan jumlah kontribusi data terhadap pembentukan klasifikasi pada *node* tersebut.

Hasil klasifikasi dari seluruh pohon menggunakan metode *XGBoost* akan digunakan untuk menentukan prediksi kelas akhir setiap observasi pada data *training*. Proses ini menghasilkan hasil klasifikasi lengkap untuk seluruh wilayah

kabupaten/kota di Indonesia tahun 2022, pada data IKG jumlah 463 data diperoleh hasil klasifikasi sebagai berikut:

Tabel 4.11 Contoh Hasil Klasifikasi *XGBoost* IKG Kabupaten/Kota Tahun 2022

Kabupaten/Kota	Aktual	Prediksi
Manokwari	Sedang	Sedang
Kota Banjar Baru	Sedang	Sedang
Kota Singkawang	Rendah	Rendah
Tapanuli Selatan	Tinggi	Tinggi
Aceh Barat	Rendah	Rendah
Yalimo	Tinggi	Tinggi
Puncak Jaya	Tinggi	Tinggi
Sampang	Tinggi	Tinggi
Kendal	Rendah	Rendah
Kota Makasar	Rendah	Rendah

Pada tabel 4.11 merupakan contoh hasil klasifikasi menggunakan metode *XGBoost* dengan data *training*, hasil klasifikasi terdapat 136 kabupaten/kota dengan kategori kelas rendah, diantaranya adalah Kota Bandung, Jombang, Ponorogo, Kota Batam, dan Kota Makasar. Wilayah dengan kategori IKG rendah menunjukkan bahwa tingkat kesejahteraan kesetaraan gender yang terjadi di tempat tersebut berada pada posisi yang baik. Kabupaten/kota dengan kategori IKG sedang berjumlah sebanyak 123, beberapa diantaranya adalah Kota Bogor, Magelang, Bangka, Lumajang, dan Probolinggo. Kabupaten/kota dengan kategori IKG tinggi berjumlah 204 kabupaten/kota, beberapa diantaranya Bengkulu utara, Sumenep, Puncak Jaya, Garut, dan Purwakarta. Daerah dengan kategori IKG tinggi memiliki arti bahwa tingkat kesetaraan gender pada wilayah tersebut masih kurang optimal, sehingga pembangunan manusia yang terjadi di daerah tersebut masih terkoreksi dengan ketimpangan gender yang cukup besar.

4.3.1 Pembentukan Pohon Klasifikasi *XGBoost*

Pembentukan pohon klasifikasi menggunakan persamaan metode *XGBoost* bertujuan untuk menunjukkan proses matematis dalam menentukan pemisahan data berdasarkan nilai *gain* pada setiap atribut. Proses ini mencakup perhitungan nilai *gradient* dan *hessian*, nilai skor *leaf*, dan perhitungan *gain* untuk memilih pembelahan terbaik. Berikut adalah pemodelan pohon menggunakan perhitungan manual dengan persamaan metode *XGBoost*:

Tabel 4.12 Contoh Data IKG Perhitungan Manual *XGBoost*

Wilayah	IKG	Kelas	Label	X_1	X_2
Pati	30,30	Rendah	0	1,30	27,50
Blitar	35,50	Rendah	0	5,40	21,40
Malang	40,20	Sedang	1	0,06	0,31
Lumajang	47,20	Sedang	1	8,60	39,20
Pamekasan	64,40	Tinggi	2	58,40	33,50
Sumenep	57,00	Tinggi	2	22,70	38,40

Tahap awal perhitungan menggunakan metode *XGBoost* yaitu inisialisasi prediksi awal, dalam kasus *multiclass classification* diasumsikan bahwa setiap kelas memiliki peluang yang sama untuk masing-masing kelas rendah, sedang, dan tinggi. Berikut perhitungan manual menggunakan persamaan metode *XGBoost*:

$$\begin{aligned}\hat{y}_i^{(0)} &= \frac{1}{N} \sum_{k=1}^N y_i \\ &= \log \left(\frac{1}{3} \right) = -1,10\end{aligned}$$

Sehingga didapatkan prediksi awal untuk semua data adalah $\hat{y}_i^{(0)} = -1,10$.

Selanjutnya menghitung menghitung *objective function* yang digunakan untuk optimasi model *XGBoost* pada klasifikasi. Sebelum menghitung *objective*

function, terlebih dahulu menghitung kemungkinan *split* data pada salah satu fitur. Dalam hal ini dipilih fitur X_1 yaitu persentase persalinan difasilitasi kesehatan, dengan melakukan kombinasi *split* sebanyak lima kali hingga dipilih nilai *gain* tertinggi.

Tabel 4.13 Perhitungan Kombinasi *Split* pada Fitur Persentase Persalinan Tidak Difasilitasi Kesehatan (X_1)

Kombinasi Split	Nilai Gain
$X_1 < 0,68$	0,60
$X_1 < 3,35$	0,07
$X_1 < 7,00$	3,00
$X_1 < 15,65$	1,50
$X_1 < 40,55$	0,60

Dari tabel 4.13 dapat dilihat bahwa *split* pada $X_1 < 7,00$ memberikan nilai *gain* tertinggi, yaitu 3,00. Oleh karena itu, *split* digunakan untuk pembentukan root node dalam memisahkan data pada pohon keputusan. Selanjutnya menghitung nilai *gradient* dan *hessian* untuk setiap data berdasarkan prediksi menggunakan *split* $X_1 < 7,00$. Nilai *gradient* dan *hessian* digunakan untuk mengetahui arah dan besarnya koreksi yang perlu dilakukan pada setiap iterasi, sehingga pembentukan *split* dapat dilakukan secara optimal untuk meningkatkan akurasi klasifikasi. *Gradient* dan *hessian* dihitung menggunakan persamaan 2.14 untuk *gradient* dan 2.15 untuk *hessian*, berikut perhitungan *gradient* dan *hessian* dengan prediksi awal $\hat{y}_i^{(0)} = -1,10$ untuk semua data. Berikut adalah perhitungan *gradient* dan *hessian* untuk masing-masing data:

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) = \hat{y}_i - y_i$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) = \hat{y}_i (1 - \hat{y}_i)$$

Tabel 4.14 Perhitungan *Gradient* dan *Hessian* pada Persentase Persalinan Tidak Difasilitasi Kesehatan (X_1)

Wilayah	Label	g_0	h_0	Kategori
Pati	0	-0,67	0,22	Left
Blitar	0	-0,67	0,22	Left
Malang	1	0,33	0,22	Left
Lumajang	1	0,33	0,22	Right
Pamekasan	2	0,33	0,22	Right
Sumenep	2	0,33	0,22	Right

Total *gradient* dan *hessian* cabang kiri:

$$G_L = \sum_{i=1}^n g_i = 0,33 - 0,67 - 0,67 = -1,00$$

$$H_L = \sum_{i=1}^n h_i = 0,248 + 0,248 = 0,496$$

Total *gradient* dan *hessian* cabang kanan:

$$G_R = \sum_{i=1}^n g_i = 0,33 + 0,33 + 0,33 = 1$$

$$H_R = \sum_{i=1}^n h_i = 0,22 + 0,22 + 0,22 = 0,66$$

Setelah menghitung nilai *gradient* dan *hessian* untuk setiap data, langkah selanjutnya adalah menghitung bobot *leaf* (w) untuk setiap simpul pada pohon. Sebelum menghitung bobot *leaf* (w) menentukan nilai lambda yang akan digunakan menggunakan kombinasi dengan memilih nilai gain tertinggi.

Tabel 4.15 Perhitungan Kombinasi Nilai *Lambda*

λ/γ	0	0,1	0,3	0,5	1
1	0,18	-0,09	-0,29	-0,49	-0,99
2	-0,05	-0,25	-0,45	-0,65	-1,15
3	-0,11	-0,31	-0,51	-0,71	-1,21
4	-0,13	-0,33	-0,53	-0,73	-1,23
5	-0,14	-0,34	-0,54	-0,74	-1,24
6	-0,14	-0,34	-0,54	-0,74	-1,24
7	-0,14	-0,34	-0,54	-0,74	-1,24
8	-0,14	-0,34	-0,54	-0,74	-1,24
9	-0,14	-0,34	-0,54	-0,74	-1,24
10	-0,14	-0,34	-0,54	-0,74	-1,24

Berdasarkan tabel 4.15 Perhitungan manual bobot *leaf* dengan mengansumsikan nilai $\lambda = 1$, pemilihan nilai $\lambda = 1$ dengan kombinasi lamda 1 sampai 10 dan dipilih nilai lamda dengan hasil gain tertinggi. Perhitungan bobot *leaf* (w) untuk setiap *node* menggunakan persamaan sebagai berikut:

$$w = -\frac{G}{H + \lambda}$$

Bobot *leaf* ini merepresentasikan kontribusi nilai prediksi pada simpul tersebut dalam mengoreksi kesalahan prediksi sebelumnya, nilai bobot *leaf* negatif menunjukkan bahwa simpul tersebut memberikan kontribus untuk menurunkan nilai prediksi dengan memperbaiki ke nilai yang lebih rendah.

Perhitungan bobot *leaf* (w) cabang kiri:

$$w = -\frac{G}{H + \lambda} = -\frac{-1}{0,66 + 1} = 0,60$$

Perhitungan bobot *leaf* (w) cabang kanan:

$$w = -\frac{G}{H + \lambda} = -\frac{1}{0,66 + 1} = -0,60$$

Langkah terakhir pada perhitungan manual menggunakan metode *XGBoost* adalah mencari nilai *gain* tertinggi pada semua variabel yang nantinya digunakan untuk menentukan pembagian dataset menjadi dua kelas. Perhitungan manual nilai *gain* menggunakan persamaan 2.21:

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma$$

Berikut perhitungan manual nilai *gain* pada indikator persalinan tidak difasilitasi kesehatan (X_1). Misalkan nilai *split* $X_1 < 7,00$ untuk membagi data menjadi dua kelas, yaitu kelas kiri (*left*) untuk nilai $X_1 < 7,00$ dan kelas kanan (*right*) untuk nilai $X_1 \geq 7,00$.

$$\begin{aligned} Gain &= \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \\ &= \frac{1}{2} \left(\frac{(-1)^2}{0,67 + 1} + \frac{(0,1)^2}{0,67 + 1} - \frac{(-1 + 1)^2}{0,67 + 0,67 + 1} \right) - 0 \\ &= \frac{1}{2} (0,60 + 0,60 - 0) \\ &= 0,60 \end{aligned}$$

Perhitungan *gain* menunjukkan bahwa pemisahan data berdasarkan fitur $X_1 < 7,00$ menghasilkan nilai *gain* sebesar 0,60. Nilai ini menunjukkan adanya peningkatan fungsi objektif yang cukup signifikan, sehingga pemilahan dianggap valid dan layak untuk diterapkan. Nilai *gain* positif menunjukkan bahwa *split* ini efektif dalam membedakan target kelas. Oleh karena itu, pembentukan struktur pohon dapat dilanjutkan dengan dua cabang (*leaf*) pada *node* kiri dan kanan, masing-masing diberi bobot sesuai hasil perhitungan.

4.3.2 Tingkat Akurasi Ketepatan Klasifikasi *XGBoost*

Tingkat akurasi dari pohon klasifikasi IKG yang dihasilkan digunakan untuk menetapkan kategori rendah, sedang, atau tinggi pada Kabupaten/Kota di Indonesia. Tingkat akurasi model dapat dievaluasi menggunakan nilai APER (apparent error rate) dan 1-APER (*total accuracy rate*). Tingkat ketepatan hasil klasifikasi IKG pada metode *XGBoost* dapat diukur menggunakan data *testing* yang ditampilkan dalam tabel *confusion matrix* berikut.

Tabel 4.16 *Confusion Matrix XGBoost*

Pengamatan	Prediksi			Total
	Rendah	Sedang	Tinggi	
Rendah	13	1	0	14
Sedang	1	9	2	12
Tinggi	0	1	24	25
Total	14	11	26	51

Hasil perhitungan *confusion matrix* berdasarkan tabel 4.16 menggunakan data *testing* IKG sebanyak 51 diperoleh 13 Kabupaten/Kota yang memiliki kategori IKG rendah diprediksi benar menjadi Kabupaten/Kota dengan kategori IKG rendah, 9 Kabupaten/Kota diprediksi benar memiliki kategori IKG sedang, dan 24 Kabupaten/Kota tepat diprediksi dengan kategori IKG tinggi. Sedangkan terdapat 5 Kabupaten/Kota yang tidak dapat diklasifikasikan dengan tepat oleh *XGBoost*, dengan rincian 1 Kabupaten/Kota dari kategori IKG rendah salah diklasifikasikan ke dalam kategori IKG sedang, 1 Kabupaten/Kota dari kategori IKG sedang, salah diklasifikasikan ke dalam kategori IKG rendah, 2 Kabupaten/Kota dari kategori IKG sedang salah diklasifikasikan ke dalam kategori IKG tinggi, dan 1 Kabupaten/Kota dari kategori IKG tinggi salah

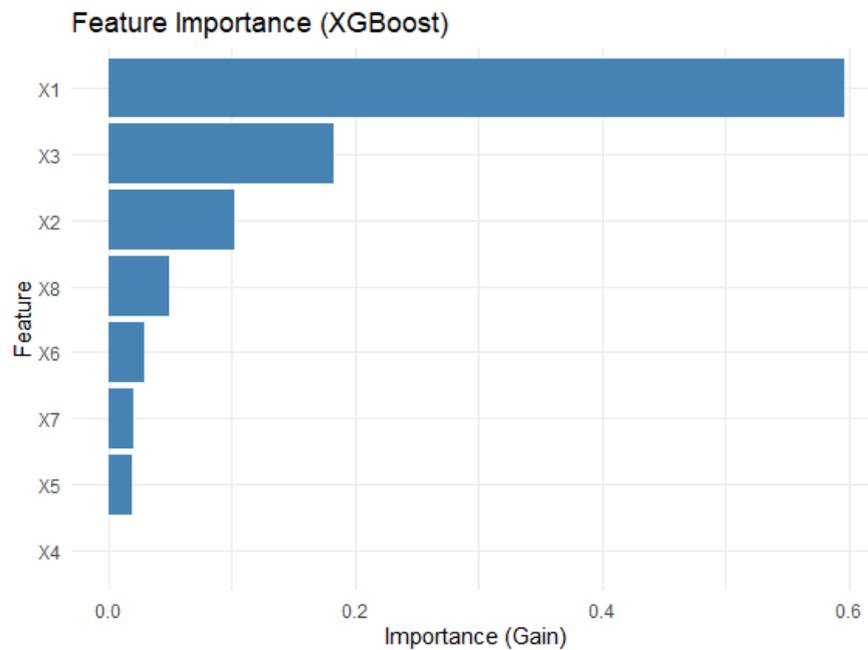
diklasifikasikan ke dalam kategori IKG sedang. Kemudian perhitungan manual pada data *testing* adalah sebagai berikut:

$$\begin{aligned}
 APER &= \frac{FP + FN}{TP + TN + FP + FN} \cdot 100\% \\
 &= \frac{1 + 0 + 2 + 1 + 0 + 1}{51} \cdot 100\% \\
 &= 9,80\% \\
 akurasi &= 100 - APER \\
 &= 90,20
 \end{aligned}$$

Nilai akurasi dari hasil klasifikasi IKG menggunakan metode *XGBoost* pada data *testing* diperoleh sebesar 90,20% dengan total pengamatan yang benar diklasifikasikan oleh model adalah 46 dari 51 pengamatan. Angka ini sudah cukup menunjukkan bahwa metode *XGBoost* dapat melakukan analisis data klasifikasi dengan baik pada data indeks ketimpangan gender Kabupaten/Kota di Indonesia tahun 2022.

4.3.3 Variabel Important XGBoost

Pada analisis klasifikasi indeks ketimpangan gender dengan metode *XGBoost* dapat diketahui variabel-variabel yang memiliki pengaruh signifikan terhadap proses klasifikasi atau juga dapat disebut sebagai *variable importance*. Semakin besar nilai *importance* suatu variabel, semakin besar kontribusinya dalam meningkatkan akurasi model klasifikasi.



Gambar 4.5 Variabel Important XGBoost

Pada Gambar 4.5 menggambarkan *variabel important* menggunakan metode *XGBoost* pada klasifikasi indeks ketimpangan gender di Indonesia. *Variabel important* diperoleh berdasarkan nilai gain, yaitu peningkatan fungsi objektif yang dihasilkan oleh variabel tersebut selama proses *split* dalam pohon keputusan. Variabel dengan nilai *gain* tertinggi menunjukkan pengaruh terbesar terhadap hasil klasifikasi. Berdasarkan diagram variabel important diatas, menunjukkan variabel X_1 yaitu indikator persalinan tidak difasilitasi kesehatan memiliki bar tertinggi dibandingkan dengan variabel lainnya. Hal ini menunjukkan bahwa variabel X_1 yaitu persalinan tidak difasilitasi kesehatan memiliki pengaruh tingkat kepentingan paling tinggi selama proses klasifikasi indeks ketimpangan gender menggunakan metode *XGBoost* dengan nilai gain sebesar 0,6. Variabel dengan nilai gain terendah adalah variabel X_4 yaitu indikator persentase perempuan yang duduk di parlemen. Nilai *gain* 0 pada variabel X_4 menunjukkan kontribusi X_4 terhadap model sangat kecil atau bahkan tidak

signifikan. Hal ini dapat diinterpretasikan bahwa dalam konteks data yang digunakan, representasi perempuan di parlemen belum cukup berperan dalam menjelaskan variasi tingkat ketimpangan gender antar wilayah. Tingginya pengaruh variabel X_1 sesuai dengan permasalahan akses terhadap fasilitas kesehatan, khususnya terkait kesehatan perempuan dalam masyarakat. Menurut data Badan Pusat Statistik tahun 2023, sebanyak 38,12% perempuan yang melakukan persalinan di rumah sakit pemerintah atau swasta, namun masih terdapat 13,26% perempuan khususnya di daerah pedesaan yang melakukan persalinan di rumah. Hal ini dapat dipengaruhi oleh peran gender dalam pengambilan keputusan, terutama dalam lingkungan pedesaan yang masih tradisional biasanya keputusan dalam perawatan kesehatan diambil oleh laki-laki. Kondisi ini menunjukkan bahwa kesenjangan gender dalam akses kesehatan masih menjadi isu penting yang perlu mendapatkan perhatian lebih dalam upaya meningkatkan indeks ketimpangan gender di Indonesia.

4.4 Kajian Keislaman dengan Hasil Penelitian

Berdasarkan hasil analisis menggunakan metode *XGBoost* ditemukan bahwa terdapat 115 Kabupaten/Kota masuk kategori tinggi, artinya masih terdapat wilayah dengan ketimpangan gender yang signifikan dalam berbagai aspek kesehatan, pemberdayaan dan pasar tenaga kerja. Hal ini sejalan dengan Qur'an Kemenag (2022), ayat yang menunjukkan konsep kesetaraan antara laki-laki dan perempuan dalam Islam terdapat dalam Surah An-Nisa (4:32):

"Dan janganlah kamu iri hati terhadap apa yang telah Allah anugerahkan kepada sebagian kamu melebihi sebagian yang lain. Bagi laki-laki ada bahagian

dari hasil usahanya, dan bagi perempuan (pun) ada bahagian dari hasil usahanya. Dan mohonlah kepada Allah sebagian dari karunia-Nya. Sesungguhnya Allah Maha Mengetahui segala sesuatu."

Ayat ini menjelaskan bahwa baik laki-laki maupun perempuan memiliki bagian dan hak yang setara dalam hal usaha dan pemberian Allah. Ayat ini mengingatkan agar manusia tidak iri hati atas pemberian Allah terhadap orang lain, karena setiap individu, baik laki-laki maupun perempuan, akan memperoleh bagian dari hasil usahanya sendiri. Ini mencerminkan konsep kesetaraan gender dalam Islam, di mana baik laki-laki maupun perempuan memperoleh nilai dan hak yang setara di hadapan Allah (Ghoffar dkk., 2001).

Berdasarkan ayat tersebut, maka dapat disimpulkan bahwasannya Islam menegaskan kesetaraan hak antara laki-laki dan perempuan dalam hal memperoleh pendidikan dan pengetahuan, hak untuk bekerja dan berpartisipasi dalam kehidupan sosial dan politik, dan hak untuk mendapatkan perlindungan hukum. Konsep kesetaraan gender dalam Islam yang menekankan pentingnya mendapatkan kesempatan yang sama untuk mengembangkan potensi dan kemampuan mereka. Oleh sebab itu, penting untuk memeningkatkan pendidikan dan pemahaman tentang konsep kesetaraan gender dalam masyarakat. Hal ini dapat mengurangi ketimpangan gender dan dapat meningkatkan kesadaran dan pemahaman tentang hak-hak laki-laki dan perempuan.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan rumusan masalah dan pembahasan yang telah dipaparkan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Hasil klasifikasi indeks ketimpangan gender di 514 Kabupaten/Kota Indonesia tahun 2022 menggunakan metode *XGBoost* menunjukkan model terbaik menggunakan perbandingan data *training* 90% dan data *testing* 10%. Pada hasil klasifikasi didapatkan 136 Kabupaten/Kota dengan kategori indeks ketimpangan gender rendah, 123 Kabupaten/Kota diklasifikasikan sebagai wilayah dengan skor indeks ketimpangan gender sedang, dan 204 Kabupaten/Kota berkategori indeks ketimpangan gender tinggi. Selain itu, diperoleh juga *variable Importance* dari beberapa indikator yang terdapat pada dataset indeks ketimpangan gender. Variabel persalinan yang tidak dilakukan di fasilitas kesehatan (X_1) menjadi variabel dengan tingkat kepentingan tertinggi. Sedangkan tujuh variabel lainnya memiliki pengaruh relatif rendah jika dibandingkan dengan variabel persalinan yang tidak dilakukan di fasilitas kesehatan (X_1) pada saat proses pembentukan pohon klasifikasi dengan menggunakan metode *XGBoost*.
2. Tingkat akurasi pada hasil klasifikasi indeks ketimpangan gender pada 514 Kabupaten/Kota Indonesia tahun 2022 menggunakan metode klasifikasi *XGBoost* pada data *testing* diperoleh nilai APER sebesar 9,80% hal ini berarti

model yang telah terbentuk mampu melakukan klasifikasi dengan tingkat keakuratan sebesar 90,20%.

5.2 Saran

Sebagai bentuk upaya dalam pengembangan serta penyempurnaan penelitian berkesinambungan. Peneliti menyarankan agar pada penelitian selanjutnya menggunakan metode klasifikasi yang berbeda sehingga dapat dilakukan perbandingan tingkat akurasi antar metode klasifikasi. Selain itu, pada penelitian selanjutnya dapat dilakukan dengan menggunakan kombinasi data *training* dan *testing* yang berbeda dalam pembagian dataset.

DAFTAR PUSTAKA

- Adi, N., Kurniawan, bayu D., Santoso, D. H., Tusianti, E., Handayani, P. W., & Karyono, Y. (2023). *Indeks Ketimpangan Gender 2022*.
- Apriliandra, S., & Krisnani, H. (2021). *Perilaku Diskriminatif pada Perempuan Akibat Kuatnya Budaya Patriarki di Indonesia Ditinjau dari Perspektif Konflik*. 3.
- Az Zuhaili, W. (2018). Tafsir Al-Munir Jilid 14 (Juz 27 & 28). *Gema Insani*, 721.
- Breiman, L., Friedman, J., Olshen, R. ., & Stone, C. J. (1984). *Classification and Regression Trees* (1stEditio).
<https://doi.org/https://doi.org/10.1201/9781315139470>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *The Journal of the Association of Physicians of India*, 42(8), 665.
- Ghoffar, M. A., Mu'thi, A., & Al-Atsari, A. I. (2001). *Tafsir Ibnu Katsir 2.3.pdf* (p. 282).
- Han, J. (2011). Data Mining Concepts and Techniques. In *Report by the Royal Society* (Issue January).
<https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf%0Ahttps://www.privacyinternational.org/node/1525%0Ahttps://ico.org.uk/media/2013894/ico-feedback-request-profiling-and-automated-decision-making.pdf>
- Haumahu, J. P., Permana, S. D. H., & Yaddarabullah, Y. (2021). Fake news classification for Indonesian news using Extreme Gradient Boosting (XGBoost). *IOP Conference Series: Materials Science and Engineering*, 1098(5), 052081. <https://doi.org/10.1088/1757-899x/1098/5/052081>
- Id, I. D. (2021). Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python. In *Ur Press* (Issue December).
<https://doi.org/10.5281/zenodo.5113507>
- Kartini, A., & Maulana, A. (2019). Redefenisi Gender dan Seks. *Jurnal Kajian Perempuan & Keislaman*, 12(2), 217–239.
- Kaur, D., Bedi, R., & Gupta, S. K. (2015). Review of Decision Tree Data mining Algorithms: ID3 and C4.5. *Proceeding of International Conference on Information Technology and Computer Science*, 8(4), 4–8.
- Lestawati, R., Rais, R., & Utami, I. T. (2018). Perbandingan Antara Metode Cart (Classification and Egression Tree) Dan Regresi Logistik (Logistic Regression) Dalam Mengklasifikasikan Pasien Penderita Dbd (Demam Berdarah Dengue). *Jurnal Ilmiah Matematika Dan Terapan*, 15(1), 98–107.
<https://doi.org/10.22487/2540766x.2018.v15.i1.10206>
- Lewis, R. J. (2000). An Introduction to Classification and Regression Tree (CART

-) Analysis. *2000 Annual Meeting of the Society for Academic Emergency Medicine*, 310, 14p.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.4103&rep=rep1&type=pdf>
- LPMQ. (2022). *Qur'an Kemenag*. Lajnah Pentashih Mushaf Al-Qur'an Kementerian Agama RI. <https://quran.kemenag.go.id/quran/per-ayat/surah/4?from=32&to=32>
- Punnoose, R., & Ajit, P. (2016). Prediction of Employee Turnover in Organizations using Machine Learning Algorithms. *International Journal of Advanced Research in Artificial Intelligence*, 5(9), 22–26.
<https://doi.org/10.14569/ijarai.2016.050904>
- Shihab, M. Q. (2006). Tafsir Al-Misbah. In *Tafsir al-Mishbah* (Vol. 14). [https://ia803106.us.archive.org/22/items/etaoin/Tafsir Al-Mishbah Jilid 14 - Dr. M. Quraish Shihab.pdf](https://ia803106.us.archive.org/22/items/etaoin/Tafsir%20Al-Mishbah%20Jilid%2014%20-%20Dr.%20M.%20Quraish%20Shihab.pdf)
- Sidiq, Y. H., & Erihadiana, M. (2022). *Gender dalam Pandangan Islam*. 5, 875–882.
- Siringoringo, R., Perangin-angin, R., & Jamaluddin. (2021). Model Hibrid Genetic-Xgboost Dan Principal Component Analysis Pada Segmentasi Dan Peramalan Pasar. *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 5(2), 97–103. <https://doi.org/10.46880/jmika.vol5no2.pp97-103>
- Taufiq, A. (2007). Islam Dan Gerakan Kesetaraan Gender Di Indonesia. *Universum*.
[http://repository.iainkediri.ac.id/id/eprint/250%0Ahttp://repository.iainkediri.ac.id/250/1/TAufiq-Islam dan Gerakan Kesetaraan Gender%2C.pdf](http://repository.iainkediri.ac.id/id/eprint/250%0Ahttp://repository.iainkediri.ac.id/250/1/TAufiq-Islam%20dan%20Gerakan%20Kesetaraan%20Gender%2C.pdf)
- Wang, T., Bian, Y., Zhang, Y., & Hou, X. (2023). Classification of Earthquakes, Explosions and Mining-induced Earthquakes Based on XGBoost Algorithm. *Computers and Geosciences*, 170(May 2022), 105242.
<https://doi.org/10.1016/j.cageo.2022.105242>
- Yolanda, A. M., Adnan, A., & Dwiputri, A. A. (2022). Pemodelan Klasifikasi Pada Indeks Ketimpangan Gender (Ikg) Tahun 2020 Dengan Metode Naïve Bayes. *Jurnal Keluarga Berencana*, 7(1), 22–31.
<https://doi.org/10.37306/kkb.v7i1.118>

LAMPIRAN

Lampiran 1 Data Indeks Ketimpangan Gender Tahun 2022

No.	Wilayah	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Kategori
1.	Simeulue	4,00	29,10	80,00	20,00	50,98	40,43	84,15	43,73	Rendah
2.	Aceh Singkil	19,10	28,30	88,00	12,00	38,35	37,36	83,76	30,33	Rendah
3.	Aceh Selatan	9,50	28,70	90,00	10,00	37,46	29,16	84,44	38,30	Rendah
4.	Aceh Tenggara	26,30	26,40	90,00	10,00	55,48	46,46	82,07	54,03	Rendah
5.	Aceh Timur	35,40	26,80	90,00	10,00	32,20	29,14	81,80	35,39	Rendah
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
169.	Cirebon	21,70	23,20	72,00	28,00	32,93	26,68	82,13	48,20	Sedang
170.	Majalengka	9,60	34,20	88,00	12,00	25,14	20,48	83,24	49,52	Sedang
171.	Sumedang	11,50	33,20	82,00	18,00	34,59	27,35	82,26	47,42	Sedang
172.	Indramayu	8,20	30,70	65,31	34,69	33,10	25,33	84,31	52,98	Sedang
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
431.	Buton Utara	26,80	35,60	75,00	25,00	43,49	34,99	91,57	71,02	Tinggi
432.	Konawe Utara	19,10	46,50	85,00	15,00	43,89	34,95	86,43	55,15	Tinggi
433.	Kolaka Timur	26,90	38,90	60,00	40,00	33,86	35,29	87,76	49,89	Tinggi
434.	Konawe Kepulauan	32,90	42,70	90,00	10,00	42,05	34,60	82,77	55,06	Tinggi
435.	Muna Barat	42,20	30,60	80,00	20,00	36,42	29,75	92,73	72,64	Tinggi
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
510.	Puncak	0,00	39,10	96,00	4,00	8,53	3,65	98,55	96,26	Rendah
511	Dogiyai	0,00	34,90	100,00	0,00	17,49	8,76	79,72	83,96	Rendah
512	Intan Jaya	42,20	45,30	100,00	0,00	21,65	6,18	88,30	71,57	Rendah
513	Deiyai	100,00	17,60	100,00	0,00	15,54	4,51	84,07	58,49	Rendah
514	Kota Jayapura	5,70	23,70	77,50	22,50	76,55	69,28	70,18	49,59	Rendah

Lampiran 2 Syntax Pembagian Data *Training* dan Data *Testing*

```

library(readxl)
library(rpart)
library(rpart.plot)
library(caret)
library(xgboost)

# Membaca data
data = read_excel("E:/SKRIPSI 4/IKG 2022.xlsx")
str(data)

# Membagi data train dan test (90:10)
set.seed(123)
index_train = sample(1:nrow(data), 0.9 * nrow(data))
train_data = data[index_train, ]
test_data = data[-index_train, ]

```

Lampiran 3 *Syntax* Pembentukan Pohon Klasifikasi CART

```
# Pohon Keputusan CART
cart_model = rpart(kelas ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8,
  data = train_data,
  method = "class",
  parms = list(split = "gini"),
  control = rpart.control(minsplit = 10, cp = 0.01, maxdepth = 5)
)
```

Lampiran 4 *Syntax* Pembentukan Pohon Klasifikasi XGBoost

```
# Data untuk XGBoost
data$Label = as.numeric(factor(data$kelas,
  levels = c("Rendah", "sedang", "Tinggi")))
fitur = as.matrix(data[, paste0("X", 1:8)])
label = data$Label
train_x = fitur[index_train, ]
train_y = label[index_train]
test_x = fitur[-index_train, ]
test_y = label[-index_train]
dtrain = xgb.DMatrix(data = train_x, label = train_y)
dtest = xgb.DMatrix(data = test_x, label = test_y)
```

Lampiran 5 *Syntax* Lanjutan Pembentukan Pohon Klasifikasi XGBoost

```
# Tuning hyperparameter XGBoost dengan Grid Search
grid = expand.grid(
  eta = c(0.01, 0.1, 0.3),
  max_depth = c(3, 5, 7),
  gamma = 0:5,
  lambda = 0:5,
  subsample = c(0.7, 1),
  nrounds = 100
)
hasil = data.frame()
for (i in seq_len(nrow(grid))) {
  p = grid[i, ]
  param = list(
    objective = "multi:softprob", num_class = 3,
    eta = p$eta, max_depth = p$max_depth,
    gamma = p$gamma, lambda = p$lambda,
    subsample = p$subsample, eval_metric = "mlogloss"
  )
  cv = xgb.cv(params = param, data = dtrain,
    nrounds = p$nrounds, nfold = 5,
    early_stopping_rounds = 10, verbose = 0)
  hasil[i, ] = c(p, logloss = min(cv$evaluation_log$test_mlogloss_mean))
}
```

Lampiran 6 Syntax Evaluasi Model XGBoost

```
# Evaluasi model XGBoost
pred_xgb = predict(model, dtest)
aktual = factor(label_test, labels = c("Rendah", "Sedang", "Tinggi"))
hasil = factor(pred_xgb, labels = c("Rendah", "Sedang", "Tinggi"))

matriks = table(Aktual = aktual, Prediksi = hasil)
print(matriks)

error_xgb = sum(matriks) - sum(diag(matriks))
APER_xgb = error_xgb / sum(matriks)
akurasi_xgb = 1 - APER_xgb
print(paste("XGBoost - APER:", round(APER_xgb, 4)))
print(paste("XGBoost - Akurasi:", round(akurasi_xgb, 4)))
```

Lampiran 7 Syntax Plot Variable Importance

```
# Importance variable XGBoost
importance = xgb.importance(feature_names = colnames(fitur), model = model)
print(importance)
xgb.plot.importance(importance)
```

Lampiran 8 Hasil Prediksi Klasifikasi Metode *Extreme Gradient Boosting* Data Training

No	Wilayah	Aktual	Prediksi
1	Luwu	Tinggi	Tinggi
2	Halmahera Barat	Tinggi	Tinggi
3	Kota Bogor	Sedang	Sedang
4	Aceh Tamiang	Sedang	Sedang
5	Magelang	Sedang	Sedang
6	Kolaka	Sedang	Sedang
7	Rote Ndao	Tinggi	Tinggi
8	Bengkulu Utara	Tinggi	Tinggi
9	Alor	Tinggi	Tinggi
10	Ponorogo	Rendah	Rendah
11	Jombang	Rendah	Rendah
21	Hulu Sungai Selatan	Sedang	Sedang
22	Tulang Bawang Barat	Sedang	Sedang
23	Manokwari	Sedang	Sedang
24	Kota Banjar Baru	Sedang	Sedang
25	Kota Singkawang	Rendah	Rendah
26	Tapanuli Selatan	Tinggi	Tinggi
27	Aceh Barat	Rendah	Rendah
28	Yalimo	Tinggi	Tinggi
29	Puncak Jaya	Tinggi	Tinggi

RIWAYAT HIDUP



Hafiz Alda Cahyafitri, lahir di Blitar pada 4 Desember 2001. Akrab dipanggil dengan Hafiz, merupakan anak ke tiga dari tigas bersaudara, pasangan bapak Kasiana dan Ibu Winarti yang bertempat tinggal di Jatinom, Kecamatan Kanigoro Kabupaten Blitar. Pendidikan formal yang ditempuh oleh peneliti yaitu pendidikan sekolah menengah atas di SMAN 1 Talun dan lulus pada tahun 2020. Peneliti melanjutkan pendidikannya di Universitas Islam Negeri Maulana Malik Ibrahim Malang pada Program Studi Matematika melalui jalur SBMPTN dengan NIM 200601110116. Untuk menambah pengalaman selama kuliah peneliti juga turut serta dalam beberapa kepanitiaan di Kampus. Peneliti menerima segala saran, kritikan, ataupun masukan demi manfaatnya tugas akhir ini. Peneliti dapat dihubungi melalui email: hafizalda01@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 : Hafiz Alda Cahyafitri
 NIM : 200601110116
 Fakultas/Jurusan : Sains dan Teknologi/Matematika
 Judul Skripsi : Implementasi Metode *Extreme Gradient Boosting* dalam
 Penentuan Klasifikasi Indeks Ketimpangan Gender di
 Indonesia
 Pembimbing I : Ria Dhea Layla Nur Karisma, M.Si.
 Pembimbing II : Evawati Alisah, M.Pd.

No	Tanggal	Hal	Tanda Tangan
1.	11 Juli 2024	Konsultasi Topik dan Data	1.
2.	18 Juli 2024	Konsultasi Bab I, II, dan III	2.
3.	25 Juli 2024	Konsultasi Kajian Agama Bab I dan II	3.
4.	5 Agustus 2024	Konsultasi Revisi Kajian Agama Bab I dan II	4.
5.	12 Agustus 2024	ACC Bab I, II, dan III	5.
6.	20 Agustus 2024	ACC Kajian Agama Bab I dan II	6.
7.	21 Agustus 2024	ACC Seminar Proposal	7.
8.	2 September 2024	Konsultasi Revisi Seminar Proposal	8.
9.	12 September 2024	Konsultasi Bab IV dan V	9.
10.	14 Oktober 2024	ACC Bab IV dan V	10.
11.	28 Oktober 2024	Konsultasi Kajian Agama Bab IV	11.
12.	13 November 2024	ACC Kajian Agama Bab IV	12.
13.	15 November 2024	ACC Seminar Hasil	13.
14.	21 November 2024	Konsultasi Revisi Seminar Hasil	14.



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

No	Tanggal	Hal	Tanda Tangan
15.	8 Mei 2025	ACC Sidang Skripsi	15.
16.	16 Juni 2025	ACC Akhir	17.

Malang, 16 Juni 2025

Mengetahui,

Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc.

NIP. 19741129 200012 2 005