

**PENENTUAN KLASIFIKASI INDEKS KETIMPANGAN
GENDER DI INDONESIA MENGGUNAKAN METODE
*RANDOM FOREST CLASSIFIER***

SKRIPSI

**OLEH
ANINDYA LUTHFIANI SUSETYO
NIM. 200601110085**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**PENENTUAN KLASIFIKASI INDEKS KETIMPANGAN
GENDER DI INDONESIA MENGGUNAKAN METODE
*RANDOM FOREST CLASSIFIER***

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
ANINDYA LUTHFIANI SUSETYO
NIM. 200601110085**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**PENENTUAN KLASIFIKASI INDEKS KETIMPANGAN
GENDER DI INDONESIA MENGGUNAKAN METODE
RANDOM FOREST CLASSIFIER**

SKRIPSI

**Oleh
Anindya Luthfiani Susetyo
NIM. 200601110085**

Telah Diperiksa dan Disetujui Untuk Diuji
Malang, 14 Juni 2024

Dosen Pembimbing I



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

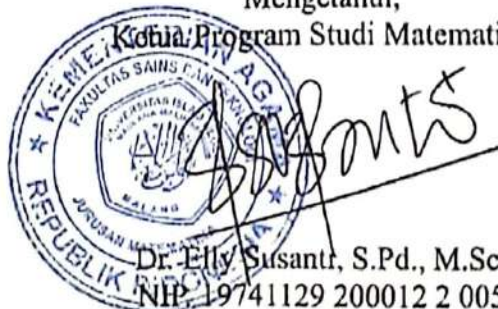
Dosen Pembimbing II



Juhari, M.Si.
NIPPPK. 19840209 202321 1 010

Mengetahui,

Kotia Program Studi Matematika



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

**PENENTUAN KLASIFIKASI INDEKS KETIMPANGAN
GENDER DI INDONESIA MENGGUNAKAN METODE
RANDOM FOREST CLASSIFIER**

SKRIPSI

Oleh
Anindya Luthfiani Susetyo
NIM. 200601110085

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)
Tanggal 21 Juni 2024

Kctua Penguji : Prof. Dr. Sri Harini, M.Si

Anggota Penguji I : Hisyam Fahmi M.Kom

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

Anggota Penguji III : Juhari, M.Si.


.....

.....

.....

.....

Mengetahui,
Ketua Program Studi Matematika


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

MOTO

“And Allah is best of planners”

(Q.S. Al-Anfal:30)

PERSEMBAHAN

Skripsi ini penulis persembahkan kepada:

Bapak dan Ibu tercinta Edi dan Tanri yang senantiasa memberikan doa, dukungan, nasihat, dan motivasi terbaik untuk kesuksesan penulis. Adik tersayang Dhimas Brillian Susetyo yang selalu memberikan doa dan semangat kepada penulis. Serta sahabat-sahabat penulis yang selalu memberikan bantuan dan semangat dalam menyelesaikan skripsi ini.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji bagi Allah SWT atas segala rahmat, taufik, serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penentuan Klasifikasi Indeks Ketimpangan Gender di Indonesia Menggunakan Metode *Random Forest Classifier*.” Shalawat dan salam semoga selalu tercurahkan kepada Nabi Muhammad *shallahu 'alaihi wa sallam* yang telah membawa manusia dari zaman jahiliah menuju ke zaman islamiah.

Ucapan syukur dan terima kasih kepada berbagai pihak yang telah memberikan bantuan, bimbingan dan arahan kepada penulis. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si, selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Elly Susanti, S.Pd., 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 memberikan berbagai pengetahuan, nasihat, motivasi dan arahan kepada penulis.
5. Juhari, M.Si., selaku selaku dosen pembimbing II yang telah memberikan bimbingan, nasihat, ilmu dan arahan kepada penulis.
6. Prof. Dr. Sri Harini, M.Si., selaku ketua penguji dalam ujian skripsi yang telah memberikan saran yang bermanfaat bagi penulis
7. Hisyam Fahmi, M.Kom., selaku penguji 1 dalam ujian skripsi yang telah memberikan saran yang bermanfaat bagi penulis.
8. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
9. Bapak Edi Susetyo dan Ibu Tanri Janah selaku orang tua penulis serta seluruh keluarga yang senantiasa mendoakan, memberikan semangat dan dukungan sehingga penulis dapat menyelesaikan tugas akhir.

10. Seluruh mahasiswa Matematika angkatan 2020 yang telah memberikan semangat, bantuan dan motivasi terbaik.

Penulis berharap adanya skripsi ini dapat memberikan manfaat dan menambah wawasan bagi penulis maupun pembaca yang tertarik dengan topik yang dibahas. Selain itu, diharapkan skripsi ini dapat menjadi bahan referensi bagi penelitian lanjutan yang lebih menyeluruh.

Wassalamu 'alaikum Warahmatullahi Wabarakatuh

Malang, 21 Juni 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR SIMBOL	xiv
DAFTAR LAMPIRAN	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	6
1.4 Manfaat Penelitian	6
1.5 Definisi Istilah	7
BAB II KAJIAN TEORI	8
2.1 <i>Synthetic Minority Over-Sampling Technique (Smote)</i>	8
2.2 <i>Random Forest</i>	9
2.2.1 <i>Out-Of-Bag Estimation</i>	13
2.2.2 <i>Bootstrap Aggregating</i>	14
2.2.3 Pembentukan Model	15
2.3 <i>K-Fold Cross Validation</i>	18
2.4 Pengukuran Tingkat Akurasi Klasifikasi	19
2.5 <i>Variable Importance</i>	20
2.6 Ketimpangan Gender	21
2.7 Indeks Khusus Ketimpangan Gender	24
2.8 Kajian Integrasi Ketimpangan Gender Dengan Al-Quran	27
2.9 Kajian Ketimpangan Gender Dengan Teori Pendukung	29
BAB III METODE PENELITIAN	31
3.1 Jenis Penelitian	31
3.2 Data Dan Sumber Data	31
3.3 Variabel Penelitian	31
3.4 Tahapan Penelitian	32
3.5 Diagram Alir Penelitian	54
BAB IV PEMBAHASAN	55
4.1 Analisis Statistik Deskriptif	55
4.2 Analisis Data	59
4.2.1 Penentuan Data <i>Training</i> Dan Data <i>Testing</i>	59
4.2.2 Balancing Data Dengan Metode Smote	59

4.2.3 Hasil Klasifikasi Menggunakan Metode <i>Random Forest</i>	64
4.2.4 Evaluasi Model Metode <i>Random Forest</i>	68
4.3 Kajian Penelitian Dalam Perspektif Islam.....	76
BAB V PENUTUP	79
5.1. Kesimpulan.....	79
5.2. Saran	79
DAFTAR PUSTAKA	81
LAMPIRAN	
RIWAYAT HIDUP	

DAFTAR TABEL

Tabel 3.1	Variabel Penelitian.....	32
Tabel 3.2	Contoh Data Perhitungan Manual	33
Tabel 3.3	Data <i>Training</i> Perhitungan Manual	34
Tabel 3.4	<i>Bootstrap</i> Pertama	34
Tabel 3.5	<i>Bins</i> Variabel <i>X1</i>	36
Tabel 3.6	<i>Entropy</i> dan <i>Gain Value X1</i>	40
Tabel 3.7	<i>Bins</i> Variabel <i>X2</i>	40
Tabel 3.8	<i>Entropy</i> dan <i>Gain Value X2</i>	44
Tabel 3.9	<i>Bins</i> Variabel <i>X5</i>	45
Tabel 3.10	<i>Entropy</i> dan <i>Gain Value X5</i>	48
Tabel 3.11	<i>Bins</i> Variabel <i>X6</i>	49
Tabel 3.12	<i>Entropy</i> dan <i>Gain Value X6</i>	52
Tabel 4.1	Analisis Deskriptif Persalinan Tidak di Fasilitas Kesehatan	56
Tabel 4.2	Analisis Deskriptif Wanita Kawin Melahirkan < 20 Tahun	57
Tabel 4.3	Analisis Deskriptif Pria dan Wanita Duduk di Parlemen	58
Tabel 4.4	Analisis Deskriptif Pria dan Wanita Pendidikan Minimal SMA ..	58
Tabel 4.6	Analisis Deskriptif TPAK Usia Kerja Laki-laki dan Perempuan ..	59
Tabel 4.7	Data <i>Training</i> dan <i>Testing</i>	60
Tabel 4.8	Frekuensi Data <i>Training</i>	61
Tabel 4.9	Hasil Klasifikasi	66
Tabel 4.10	<i>K-Fold Cross Validation</i>	68
Tabel 4.11	<i>Confusion Matrix Data Testing</i>	69
Tabel 4.12	Tingkat Ketepatan Model	73

DAFTAR GAMBAR

Gambar 2.1	Proses <i>Random Forest</i>	15
Gambar 2.2	Ilustrasi Pohon <i>Random Forest</i>	18
Gambar 2.3	Skema <i>K-Fold Cross Validation</i>	18
Gambar 3.1	Diagram Alir Penelitian.....	54
Gambar 4.1	<i>Pie Chart</i> IKG	55
Gambar 4.2	Jumlah Kelas Sebelum SMOTE.....	61
Gambar 4.3	Jumlah Kelas Setelah SMOTE	62
Gambar 4.4	Grafik Hasil Percobaan <i>ntree</i>	64
Gambar 4.5	Grafik Hasil Klasifikasi Indeks Ketimpangan Gender.....	68
Gambar 4.6	<i>Variable Importance</i> IKG.....	74

DAFTAR SIMBOL

Y	: Data Indeks Ketimpangan Gender
X_1	: Data Proporsi Persalinan Tidak di Fasilitas Kesehatan
X_2	: Data Proporsi Wanita Berusia 15 – 49 Tahun yang Pernah Kawin dan Saat Melahirkan Hidup Pertama Berusia Kurang dari 20 Tahun
X_3	: Persentase Laki-Laki dan Perempuan yang Duduk di Parlemen
X_4	: Persentase Penduduk Laki-Laki dan Perempuan dengan Pendidikan Minimal SMA
X_5	: Persentase Angkatan Kerja Terhadap Penduduk Usia Kerja
Θ_k	: Pohon dengan Vektor Random Independen, Distribusi yang Sama, dan Berasal dari Komponen Terbaik
PE^*	: <i>Generalization Error Random Forest</i>
$\bar{\rho}$: Nilai Korelasi
E	: Nilai Ekspektasi
sd	: Nilai Standar Deviasi
k	: Jumlah Lipatan (<i>Fold</i>)
$\hat{\theta}^*$: Nilai Estimasi <i>Bootstrap</i>
$h_K(x)$: Sekumpulan Pohon
M	: Jumlah Variabel Prediktor
X^{*1}	: Sampel <i>Bootstrap</i> Pertama
\hat{F}^*	: Estimator <i>Bootstrap</i> untuk $\hat{\theta}$
G_F	: Indeks Perempuan
G_M	: Indeks Laki-laki
Faskes	: Proporsi Persalinan yang Tidak di Fasilitas Kesehatan
ULP	: Proporsi 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 Pendidikan Minimal SMA
SE_M	: Persentase Laki-Laki 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
Kesehatan	: Indeks Kesehatan Reproduksi
Pemberdayaan	: Indeks Pemberdayaan
IPK	: Indeks Pasar Kerja
$G_{F,M}$: Indeks Dimensi

DAFTAR LAMPIRAN

Lampiran 1	Data Indeks Ketimpangan Gender.....	48
Lampiran 2	Syntax Model <i>Random Forest</i> dengan RStudio	49
Lampiran 3	Syntax Evaluasi Model <i>Random Forest</i> dengan RStudio.....	73
Lampiran 4	<i>Syntax Plot Variabel Importance</i>	53
Lampiran 5	<i>Output</i> Tingkat Akurasi Model <i>Random Forest</i>	54
Lampiran 6	<i>Output</i> Evaluasi Model <i>Random Forest</i>	75

ABSTRAK

Susetyo, Anindya Luthfiani. 2024. **Penentuan Klasifikasi Indeks Ketimpangan Gender di Indonesia Menggunakan Metode *Random Forest Classifier***. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Juhari, M.Si.

Kata Kunci: Ketimpangan Gender, Indeks Ketimpangan Gender, *Random Forest*, *Ensemble Learning*

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 mendapatkan klasifikasi indeks ketimpangan gender di Kabupaten/Kota Indonesia dan tingkat akurasi klasifikasi indeks ketimpangan gender dengan metode *Random Forest*. Hasil penelitian menunjukkan model terbaik klasifikasi indeks ketimpangan gender menggunakan parameter jumlah pohon atau $n_{tree} = 500$ dan $m_{try} = 2$. Hasil klasifikasi menunjukkan terdapat 219 Kabupaten/Kota dengan kategori indeks ketimpangan gender rendah, 69 Kabupaten/Kota dengan skor indeks ketimpangan gender menengah bawah, 60 Kabupaten/Kota dengan skor indeks ketimpangan gender menengah atas, dan 115 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 f_1 -score dari model klasifikasi sebesar 91,76% menunjukkan metode *Random Forest* mampu mengklasifikasikan indeks ketimpangan gender Kabupaten/Kota di Indonesia dengan sangat baik.

ABSTRACT

Susetyo, Anindya Luthfiani. 2024. **Determining the Classification of Gender Inequality Index in Indonesia Using Random Forest Classifier Method.** Thesis. Mathematics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim. Advisors: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Juhari, M.Si

Keywords: Gender Inequality, Gender Inequality Index, Random Forest, Ensemble Learning

Gender inequality is a significant social issue in various regions, including Indonesia. Measuring and classifying the gender inequality index can help the government formulate more effective policies to address this issue. This study aims to obtain the classification of the gender inequality index in regencies/cities in Indonesia and the accuracy level of gender inequality index classification using the Random Forest method. The classification results show that the best model for classifying the gender inequality index uses the parameters $n_{tree} = 500$ and $m_{try} = 2$. The classification results indicate that there are 219 regencies/cities with a low gender inequality index, 69 regencies/cities classified with a lower-middle gender inequality index score, 60 regencies/cities classified as regions with an upper-middle gender inequality index score, and 115 regencies/cities with a high gender inequality index. The variable with the most significant level of importance is childbirth not conducted in healthcare facilities (X_1). The $f1 - score$ of 91,76% from the classification model shows that the Random Forest method can classify the gender inequality index in regencies/cities in Indonesia very well.

مستخلص البحث

سوسيتيو، أنبنديا لطفباني ٢٠٢٤. تحديد تصنيف مؤشر عدم المساواة بين الجنسين في إندونيسيا، باستخدام طريقة تصنيف الغابات العشوائية. البحث العملي قسم الرياضيات، كلية العلوم والتكنولوجيا. جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرفة: (١) ريا ديا ليلي نور كرسما، الماجستير. (٢) جوهرى، الماجستير.

الكلمات المفتاحية: عدم المساواة بين الجنسين، مؤشر عدم المساواة بين الجنسين، الغابة العشوائية، التعلم القابل للإندماج

يعد عدم المساواة بين الجنسين قضية اجتماعية مهمة في مختلف المناطق، في إندونيسيا. يمكن لقياس مؤشر عدم المساواة بين الجنسين وتصنيفه أن يساعد الحكومة في صياغة سياسات أكثر فعالية لمعالجة هذه المشكلة. تهدف هذه الدراسة إلى الحصول على تصنيف مؤشر عدم المساواة بين الجنسين في المناطق أو المدن بإندونيسيا ودقة تصنيف مؤشر عدم المساواة بين الجنسين باستخدام طريقة الغابة العشوائية. وتظهر نتائج التصنيف أن هناك ٢١٩ المناطق أو المدن لها مؤشر منخفض لعدم المساواة بين الجنسين، و ٦٩ المناطق أو المدن لها درجة مؤشر متوسط أدنى لمؤشر عدم المساواة بين الجنسين، و ٦٠ المناطق أو المدن لها درجة مؤشر متوسط أعلى لمؤشر عدم المساواة بين الجنسين، و ١١٥ المناطق أو المدن لها درجة مؤشر مرتفع لعدم المساواة بين الجنسين. المتغير الأهمية الأكثر تأثيراً هو الولادة لنموذج التصنيف البالغة ٩١,٧٦٪. أن طريقة الغابة (f1). تُظهر نتيجة درجة X_1 التي لا تتم في مرفق صحي (العشوائية قادرة على تصنيف مؤشر عدم المساواة بين الجنسين في المناطق أو المدن في إندونيسيا بشكل جيد للغاية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kesetaraan gender adalah kondisi di mana laki-laki dan perempuan dapat menikmati perlakuan yang setara dan memiliki kontrol penuh atas dirinya dengan bobot yang sama untuk merealisasikan hak asasi yang dimiliki sebagai manusia serta potensinya dalam segala aspek bidang dalam kegiatan sehari-hari seperti kegiatan ekonomi, politik dan sosial budaya (KEMENPPPA, 2024). Isu kesetaraan gender masih menjadi sebuah perbincangan yang polemik di Indonesia bahkan dunia internasional. Perserikatan Bangsa-Bangsa (PBB) mencetuskan indikator yang digunakan sebagai tolak ukur untuk tingkat ketidaksetaraan gender yang terjadi di setiap negara di seluruh dunia yaitu Indeks Ketimpangan Gender (IKG). IKG terbentuk dari tiga dimensi, yaitu kesehatan, pemberdayaan, dan pasar tenaga kerja. Tujuan dari IKG adalah untuk memberikan gambaran terkait potensi yang tidak tercapai dari sektor pembangunan manusia akibat ketimpangan gender (BPS, 2022).

Perolehan nilai IKG yang semakin kecil akan menunjukkan kondisi ketimpangan gender yang terjadi di negara tersebut semakin membaik. Berdasarkan data IKG secara global tahun 2021 oleh Badan Pusat Statistik (BPS) Indonesia, diketahui wilayah dengan IKG yang rendah berada di wilayah Eropa dan Asia Tengah sebesar 0,227%. Sementara itu, di beberapa wilayah seperti negara Arab, Asia Selatan dan Sub-Sahara Afrika masih memiliki nilai IKG di atas 0,5% yang berarti bahwa kondisi ketimpangan gender di kawasan tersebut masih relatif tinggi

sehingga potensi pembangunan manusia yang dicapai terpengaruh akibat adanya ketimpangan gender. Pada tahun 2020 Indonesia memiliki nilai sebesar 0,447%, di mana ini memiliki nilai yang lebih rendah dibanding dengan angka global. Tahun 2021, Indonesia memiliki nilai IKG mencapai 0,444% dan indeks global menunjukkan nilai 0,465%. Hingga tahun 2022, ketimpangan gender di Indonesia berada diperingkat 110 dari 170 negara yang mana terjadi kenaikan 11 peringkat sejak tahun 2019. Pencapaian ini menunjukkan adanya keberhasilan upaya pemerintah Indonesia dalam melakukan mengurangi ketimpangan gender di lingkup tiga dimensi (BPS, 2022). Sektor ekonomi, pertumbuhan perekonomian suatu negara tidak hanya didukung oleh tingkat keberhasilan potensi yang dilakukan oleh pria, tetapi juga para wanita (Vera, 2016). Studi menunjukkan ketika rasio tingkat partisipasi angkatan kerja dan rasio partisipasi kasar perempuan terhadap laki-laki memiliki kenaikan, maka perekonomian juga akan mengalami pertumbuhan (Deris dkk., 2022). Selain itu, ketimpangan gender yang terjadi dalam sektor pendidikan dan kesehatan juga memiliki peran dalam pertumbuhan ekonomi. Ketika rasio pendidikan yaitu rata-rata lama sekolah antara pria dan wanita meningkat akan menyebabkan produk domestik regional bruto mengalami kenaikan dan begitu pula dengan sektor kesehatan (Sari dkk., 2019).

Selain keuntungan untuk sebuah negara dalam banyak aspek, manfaat yang didapatkan apabila tercapainya kesetaraan gender antara laki-laki dan perempuan juga disebutkan dalam Al-Quran. Melakukan pekerjaan sebenarnya sudah diungkapkan dalam Al-Qur'an, yaitu pada QS. At-Taubah:105 yang memiliki arti sebagai berikut (Kemenag, 2024).

“Katakanlah (Nabi Muhammad), “Bekerjalah! Maka, Allah, rasul-Nya, dan orang-orang mukmin akan melihat pekerjaanmu. Kamu akan dikembalikan kepada (Zat) yang mengetahui yang gaib dan yang nyata. Lalu, Dia akan memberitakan kepada kamu apa yang selama ini kamu kerjakan.”

Makna yang terkandung dalam ayat ini adalah baik laki-laki maupun perempuan memiliki tingkat yang sama dengan potensi yang setara untuk melakukan perbuatan kebaikan selama mereka memiliki keyakinan dan iman. Bagi laki-laki dan perempuan yang beriman serta konsisten dalam kebaikan, Allah SWT akan memberikan kehidupan yang baik untuk keduanya tanpa ada perbedaan bagi masing-masing gender.

Berbagai upaya dilakukan oleh Pemerintah Indonesia sebagai bentuk keseriusan dalam menanggapi isu kesetaraan gender, salah satunya tercermin dalam Instruksi Presiden Republik Indonesia Nomor 9 Tahun 2000 tentang Pengarusutamaan Gender dalam Pembangunan Nasional. Pengarusutamaan gender (PUG) merupakan strategi yang dibuat dengan tujuan terlaksananya perencanaan, penyusunan, pelaksanaan hingga evaluasi hasil atas program dan kebijakan dari pembangunan nasional dengan perspektif gender dalam rangka agar terwujudnya kesetaraan gender dalam kehidupan sehari-hari masyarakat Indonesia (KPPPA, 2022).

Klasifikasi merupakan salah satu metode dalam statistika yang digunakan untuk melakukan pengelompokan terhadap benda-benda berdasarkan karakteristik yang dimiliki (Wibawa dkk., 2018). Klasifikasi dilakukan untuk mencari suatu pola, fungsi atau model tertentu dalam data sehingga data dapat dikelompokkan menjadi kelas atau label sesuai dengan ciri-ciri atau karakteristik dari data. Terdapat banyak metode klasifikasi yang digunakan untuk analisis data antara lain *support vector machine* (SVM), *AdaBoost*, *decision tree*, dan *random forest*. Salah satu metode

klasifikasi yang populer digunakan karena memiliki tingkat akurasi yang paling baik di antara metode-metode klasifikasi yang lain adalah metode *random forest* (Alonso dkk., 2022).

Random forest merupakan metode klasifikasi yang menggunakan teknik penggabungan (*ensemble method*) yang dikembangkan oleh Breiman pada tahun 2001 (Karisma dkk., 2015). *Ensemble method* digunakan untuk meningkatkan tingkat akurasi dari klasifikasi (Parmar dkk., 2019). *Random forest* adalah metode klasifikasi yang terdiri atas kumpulan *decision tree* atau pohon keputusan yang dihasilkan dengan melakukan pemilihan subset dari *dataset* dan memilih subset variabel prediktor yang dilakukan secara acak yang kemudian hasil dari proses tersebut digabung untuk mendapatkan *random forest* (Alonso dkk., 2022). Pohon-pohon tersebut diperoleh dari sampel *bootstrap* dan dilakukan *majority vote* untuk mendapatkan hasil akhir (Alita & Isnain, 2020).

Salah satu keunggulan yang utama dari metode *random forest* adalah metode ini dapat melakukan pengukuran terhadap nilai kepentingan dari setiap data untuk mempelajari dampak yang ditimbulkan dari setiap data terhadap hasil dari prediksi kelas (Uddin & Uddin, 2015). Kelebihan metode *random forest* yang lain diantaranya adalah dapat digunakan pada data yang besar dengan tingkat akurasi tinggi, dapat memberikan estimasi variabel yang penting pada klasifikasi, dan dapat digunakan untuk menstabilkan bias pada populasi yang tidak seimbang di *dataset* dan memiliki tingkat akurasi yang lebih baik jika dibandingkan dengan metode klasifikasi lainnya (Karisma dkk., 2015).

Penelitian terkait klasifikasi terhadap faktor beresiko yang diubah dari pasien iskemik dan pasien hemoragik dengan menggunakan metode *random forest* di

dalamnya pernah dilakukan oleh Karisma dkk. (2015). Penelitian ini menggunakan proporsi data *training* dan data *testing* masing-masing 85% dan 15%. Hasil penelitian yang diperoleh menunjukkan tingkat akurasi sebesar 98,32%. Penelitian klasifikasi lainnya juga pernah dilakukan oleh Alonso dkk., (2022) dengan menggunakan algoritma klasifikasi *decision tree*, *adaboost*, *random forest*, *naïve bayes*, SVM, dan K-NN dalam memprediksi pasien rawat inap dengan gangguan kejiwaan skizofrenia. Hasil penelitian menunjukkan hasil akurasi terbaik diperoleh metode *random forest* dengan tingkat akurasi sebesar 72,7%. Kemudian, pada penelitian mengenai komparasi tiga algoritma klasifikasi, yaitu *random forest*, *adaboost*, dan *neural network* dalam analisis ujaran kebencian dan kalimat kasar yang dilakukan oleh Nugroho dkk., (2019) diperoleh tingkat akurasi terbaik pada *random forest* yaitu 0,722, sedangkan metode *adaboost* dan *neural network* secara berturut-turut memiliki skor akurasi sebesar 0,708 dan 0,596.

Berdasarkan latar belakang, maka penelitian ini memanfaatkan metode *Random Forest* dalam menentukan klasifikasi indeks ketimpangan gender di Indonesia. Penentuan faktor yang berpengaruh atau variabel terhadap indeks ketimpangan gender diperoleh dari hasil sensus yang dilakukan oleh BPS pada tahun 2022. Penelitian ini diharapkan dapat memberikan informasi kepada instansi terkait sebagai bahan pertimbangan dalam menentukan langkah dan kebijakan dalam isu ketimpangan gender di Indonesia.

1.2 Rumusan Masalah

Berdasarkan pemaparan latar belakang, maka rumusan masalah penelitian disusun sebagai berikut:

1. Bagaimana hasil klasifikasi indeks ketimpangan gender di Kabupaten/Kota Indonesia dengan metode *Random Forest*?
2. Bagaimana tingkat akurasi yang diperoleh pada klasifikasi indeks ketimpangan gender menggunakan metode *Random Forest*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dalam penelitian ini adalah sebagai berikut:

1. Mendeskripsikan hasil klasifikasi indeks ketimpangan gender di Indonesia menggunakan metode *Random Forest*.
2. Menganalisis tingkat akurasi pada klasifikasi indeks ketimpangan gender menggunakan metode *Random Forest*.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian diharapkan beberapa manfaat sebagai berikut:

1. Bagi Penulis

Manfaat bagi penulis yang diperoleh dari penelitian ini adalah dapat memperdalam wawasan dan pengetahuan peneliti mengenai penerapan metode *Random Forest* terhadap Indeks Ketimpangan Gender.

2. Bagi Instansi

Memberikan gambaran hasil terkait skor indeks ketimpangan gender di Indonesia sehingga dapat dijadikan bahan pertimbangan untuk mengurangi ketimpangan gender di Indonesia.

3. Bagi Pembaca

Memberikan informasi terkait penerapan metode *Random Forest* sebagai metode klasifikasi pada data indeks ketimpangan gender.

1.5 Definisi Istilah

- Out of Bags* : Elemen-elemen yang tidak terpakai pada data *bootstrap*.
- Margin* : Fungsi yang berguna untuk melakukan pengukuran terhadap jumlah rata-rata *vote X* dan *Y* untuk kelas yang tepat dari kelas lainnya.
- Confusion Matrix* : Alat dalam analisis data klasifikasi yang digunakan untuk melakukan evaluasi hasil dari model klasifikasi yang telah dilakukan proses analisis.
- Ketimpangan Gender : Fenomena di mana posisi perempuan memiliki tempat yang lebih rendah dibandingkan dengan laki-laki dalam banyak sektor di kehidupan sehari-hari.
- Data Training* : Data yang digunakan untuk melakukan pembentukan model klasifikasi.
- Data Testing* : Data yang digunakan untuk melakukan pengujian kebaikan model yang telah dibentuk.

BAB II

KAJIAN TEORI

2.1 *Synthetic Minority Over-Sampling Technique (SMOTE)*

Class imbalanced atau ketidakseimbangan kelas adalah kondisi di mana distribusi jumlah sampel dalam suatu *dataset* di antara kelas-kelas yang berbeda memiliki perbedaan yang drastis. Antara kelas satu atau beberapa kelas lainnya memiliki jumlah sampel yang jauh lebih banyak atau jauh lebih sedikit (Syukron dkk., 2023). Kondisi *class imbalance* pada data dapat berakibat pada hasil akhir klasifikasi tidak sesuai karena model biasanya akan cenderung untuk melakukan kategorisasi terhadap kelas mayoritas. Salah satu cara untuk mengatasi permasalahan *class imbalanced* dapat dilakukan dengan menerapkan teknik *oversampling* pada *dataset*.

Teknik *oversampling* merupakan salah satu metode *resampling* yang digunakan untuk mengatasi *class imbalance*. Pada teknik ini, sampel baru akan ditambahkan pada kelas minoritas dengan tujuan untuk menyeimbangkan kelas. Salah satu metode dalam *oversampling* yang umum digunakan adalah *synthetic minority over-sampling technique (SMOTE)*. Metode ini pertama kali dipublikasikan oleh Chawla dkk. (2002). SMOTE bekerja dengan menggunakan pendekatan replikasi terhadap kelas minoritas dengan melakukan sintesis data berdasarkan nilai *k-nearest neighbors* untuk setiap data pada kelas minoritas. Jarak *euclidian* yang mengukur kedekatan antara titik data dalam ruang digunakan untuk menentukan *nearest neighbor*. Menggunakan nilai *k-nearest neighbors* daripada melakukan duplikasi terhadap kelas minoritas untuk membentuk data sintesis

secara acak bertujuan untuk menghindari *overfitting* (Damayanti & Purwinarko, 2024).

Misalkan terdapat data dengan y variabel yaitu $\mathbf{x}^R = [x_1, x_2, \dots, x_3]$ dan $\mathbf{z}^R = [z_1, z_2, \dots, z_3]$, maka jarak *euclidian* $d(x, z)$ adalah sebagai berikut.

$$d(x, z) = \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2 + \dots + (x_y - z_y)^2} \quad (2.1)$$

Secara umum, formula untuk menentukan data sintesis adalah sebagai berikut.

$$x_{syn} = x_i + (x_{knn} - x_i) \times \delta \quad (2.2)$$

dengan

- x_{syn} : Data Sintesis Hasil Replikasi
- x_i : Data yang Akan dilakukan Replikasi
- x_{knn} : Data dengan Jarak Terdekat dari Data yang Akan dilakukan Replikasi
- δ : Bilangan Acak Antara 0 dan 1

2.2 *Random Forest*

Random forest adalah metode klasifikasi yang dikenalkan oleh Breiman pada tahun 2001. *Random forest* merupakan metode yang terbentuk atas sekumpulan struktur pohon (*trees*) $\{h(\mathbf{x}, \theta_k), k = 1, \dots\}$ yang mana θ_k adalah setiap pohon merupakan sebuah vektor random independen dengan distribusi yang sama dan setiap pohon berasal dari komponen yang terbaik (Breiman, 2001).

Misalkan terdapat *ensemble of classifiers* $h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_K(\mathbf{x})$ dengan *data training* yang dipilih secara acak dari distribusi vektor acak \mathbf{X}, Y maka dapat didefinisikan fungsi margin yang berguna untuk melakukan pengukuran terhadap jumlah rata-rata *vote* \mathbf{X} dan Y untuk kelas yang tepat dari kelas lainnya. Fungsi

margin memiliki nilai yang berbanding lurus dengan *confidence level* dalam analisis klasifikasi. Apabila margin memiliki ukuran yang besar, maka *confidence level* juga memiliki nilai yang tinggi (Kulkarni, 2013). Adapun persamaan fungsi margin adalah sebagai berikut dengan $h_K(\mathbf{x}) = h(\mathbf{X}, \theta_k)$ dalam *random forest*:

$$mr(\mathbf{X}, Y) = P_{\theta}(h(\mathbf{X}, \theta) = Y) - \max_{J \neq Y} P_{\theta}(h(\mathbf{X}, \theta) = J) \quad (2.3)$$

di mana:

- $mr(\mathbf{X}, Y)$: Fungsi Margin
- $P_{\theta}(h(\mathbf{X}, \theta) = Y)$: Peluang X dikelompokkan ke Kelas Y dengan Menggunakan Model $h(\mathbf{X}, \theta)$
- $\max_{J \neq Y} P_{\theta}(h(\mathbf{X}, \theta) = J)$: Peluang Maksimum X dikelompokkan ke Kelas Lain Selain Y dengan Menggunakan Model $h(\mathbf{X}, \theta)$

Selain itu, terdapat *generalization error* (*gen error*) dalam *random forest*. *Generalization error* merupakan jarak yang diharapkan antara *data testing* dan *data training* yang digunakan untuk melakukan pengukuran terhadap algoritma yang digunakan *overfitting* terhadap *data training* atau tidak (He dkk., 2022). *Gen error* dalam *random forest* dinotasikan dengan PE^* . Ketika jumlah dari pohon meningkat, hampir semua θ_1, \dots, PE^* akan konvergen seperti pada persamaan berikut:

$$P_{X,Y} = (P_{\theta}(h(\mathbf{X}, \theta) = Y) - \max_{J \neq Y} P_{\theta}(h(\mathbf{X}, \theta) = J) < 0) \quad (2.4)$$

Sehingga didapatkan fungsi dari *gen error* metode *random forest* adalah sebagai berikut.

$$PE^* = P_{X,Y}(mr(X, Y) < 0) \quad (2.5)$$

di mana:

- $P_{X,Y}$: Probabilitas Joint dari X dan Y
- $P_{\theta}(h(\mathbf{X}, \theta) = Y)$: Peluang X dikelompokkan ke Kelas Y dengan Menggunakan Model $h(\mathbf{X}, \theta)$
- $\max_{J \neq Y} P_{\theta}(h(\mathbf{X}, \theta) = J)$: Peluang Maksimum X dikelompokkan ke Kelas Lain Selain Y dengan Menggunakan Model $h(\mathbf{X}, \theta)$
- PE^* : *Generalization Error*
- $P_{X,Y}(mr(X, Y) < 0)$: Peluang Data X dan Y Terjadi Bersama-sama dengan Fungsi Margin $mr(X, Y) < 0$

Kemudian terdapat nilai *strength* yang digunakan untuk melihat nilai rata-rata akurasi pada pohon tunggal. Nilai *strength* yang besar menunjukkan bahwa nilai akurasi dari prediksi memiliki nilai yang tinggi. Peningkatan nilai akurasi dibutuhkan dengan meminimalkan nilai korelasi ($\bar{\rho}$) dan mempertahankan nilai *strength* (s) (Breiman, 2001). Persamaan nilai *strength*, yaitu:

$$s = E_{X,Y}mr(X, Y) \quad (2.6)$$

dengan asumsi $s \geq 0$ dan diberikan pertidaksamaan *Chebychev*, yaitu

$$PE^* \leq \frac{var(mr)}{s^2}$$

misalkan

$$\hat{f}(\mathbf{X}, Y) = \arg \max_{J \neq Y} P_{\theta}(h(\mathbf{X}, \theta) = J) \quad (2.7)$$

dan

$$rmg(\theta, \mathbf{X}, Y) = I(h(\mathbf{X}, \theta) = Y) - I(h(\mathbf{X}, \theta) = \hat{f}(\mathbf{X}, Y))$$

maka dapat ditulis:

$$\begin{aligned} mr(\mathbf{X}, Y) &= P_{\theta}(h(\mathbf{X}, \theta) = Y) - P_{\theta}(h(\mathbf{X}, \theta) = \hat{f}(\mathbf{X}, Y)) \\ &= E_{\theta} \left[I(h(\mathbf{X}, \theta) = Y) - I(h(\mathbf{X}, \theta) = \hat{f}(\mathbf{X}, Y)) \right] \end{aligned} \quad (2.8)$$

dengan demikian, $mr(\mathbf{X}, Y)$ merupakan ekspektasi dari $rmg(\theta, \mathbf{X}, Y)$ terhadap θ .

Untuk semua fungsi identitas f , yaitu.

$$[E_{\theta} f(\theta)]^2 = E_{\theta, \theta'} f(\theta) f(\theta') \quad (2.9)$$

di mana θ, θ' adalah independen dengan distribusi yang sama, maka dapat dituliskan sebagai berikut:

$$mr(\mathbf{X}, Y)^2 = E_{\theta, \theta'} rmg(\theta, \mathbf{X}, Y) rmg(\theta', \mathbf{X}, Y) \quad (2.10)$$

dengan menggunakan persamaan (2.8), dapat dicari nilai dari varians *function margin* ($var(mr)$), yaitu:

$$\begin{aligned} var(mr) &= E_{\theta, \theta'} mr(\mathbf{X}, Y)^2 \\ &= E_{\theta, \theta'} rmg(\theta, \mathbf{X}, Y) rmg(\theta', \mathbf{X}, Y) \\ &= E_{\theta, \theta'} \left(cov_{\mathbf{X}, Y} rmg(\theta, \mathbf{X}, Y) rmg(\theta', \mathbf{X}, Y) \right) \\ &= E_{\theta, \theta'} \left(\rho(\theta, \theta') sd(\theta) sd(\theta') \right) \\ &= \bar{\rho} (E_{\theta} sd(\theta))^2 \end{aligned} \quad (2.11)$$

menggunakan persamaan (2.9) dengan $\rho(\theta, \theta')$ merupakan korelasi antara $rmg(\theta, \mathbf{X}, Y)$ dan $rmg(\theta', \mathbf{X}, Y)$, dapat dicari nilai korelasi $\bar{\rho}$. Persamaan nilai $\bar{\rho}$ adalah sebagai berikut:

$$\begin{aligned} var(mr) &= \bar{\rho} (E_{\theta} sd(\theta))^2 \\ \bar{\rho} &= \frac{var(mr)}{(E_{\theta} sd(\theta))^2} \\ \bar{\rho} &= \frac{E_{\theta, \theta'} \left(\rho(\theta, \theta') sd(\theta) sd(\theta') \right)}{(E_{\theta} sd(\theta)) \cdot (E_{\theta} sd(\theta))} \end{aligned}$$

$$\bar{\rho} = \frac{E_{\theta, \theta'}(\rho(\theta, \theta')sd(\theta)sd(\theta'))}{E_{\theta, \theta'}(sd(\theta)sd(\theta'))} \quad (2.12)$$

di mana:

s	:	Nilai <i>Strength</i>
$E_{X,Y}mr(X, Y)$:	Ekspektasi dari Fungsi Margin
$var(mr)$:	Varians Fungsi Margin
s^2	:	Varians
$\hat{f}(X, Y)$:	Prediksi Variabel Target (Y) Berdasarkan Variabel Input (X)
$\arg \max_{J \neq Y} P_{\theta}(h(X, \theta) = J)$:	Argumen Maksimum dari P_{θ}
$rmg(\theta, X, Y)$:	Fungsi <i>Raw</i> Margin
$[E_{\theta}f(\theta)]^2$:	Kuadrat dari Ekspektasi Fungsi $f(\theta)$ terhadap θ
$E_{\theta, \theta'}rmg(\theta, X, Y)$:	Ekspektasi Ganda dari Fungsi $rmg(\theta, X, Y)$ Terhadap θ dan θ'
$\bar{\rho}$:	Nilai Korelasi

2.2.1 *Out-of-Bag Estimation*

Setiap data *training* baru yang dihasilkan dari proses subset *bootstrap*, terdapat sepertiga dari *random sample* yang tidak digunakan karena hanya dua pertiga dari setiap pohon *data training* yang digunakan. Elemen-elemen yang tidak terpakai disebut dengan *out-of-bag* (OOB) (Louppe, 2014).

Sampel OOB dapat digunakan untuk melakukan perkiraan tingkat kepentingan (*importance value*) dari setiap variabel dan dapat digunakan untuk melakukan pengukuran validasi. Masing-masing sampel OOB untuk setiap pohon

dilakukan perhitungan akurasi dengan mengganti variabel secara acak diantara sampel lalu melakukan perhitungan ulang terhadap akurasi pada set yang baru. Estimasi OOB dapat memberikan estimasi yang akurat terhadap *generalization error* dari *ensemble* yang bahkan lebih baik daripada estimasi *K-Fold Cross Validation* (Louppe, 2014). Estimasi *K-Fold Cross Validation* menggunakan $k = 5$ atau $k = 10$ sangat direkomendasikan karena akan mengurangi bias yang ditimbulkan (Rodríguez dkk., 2010).

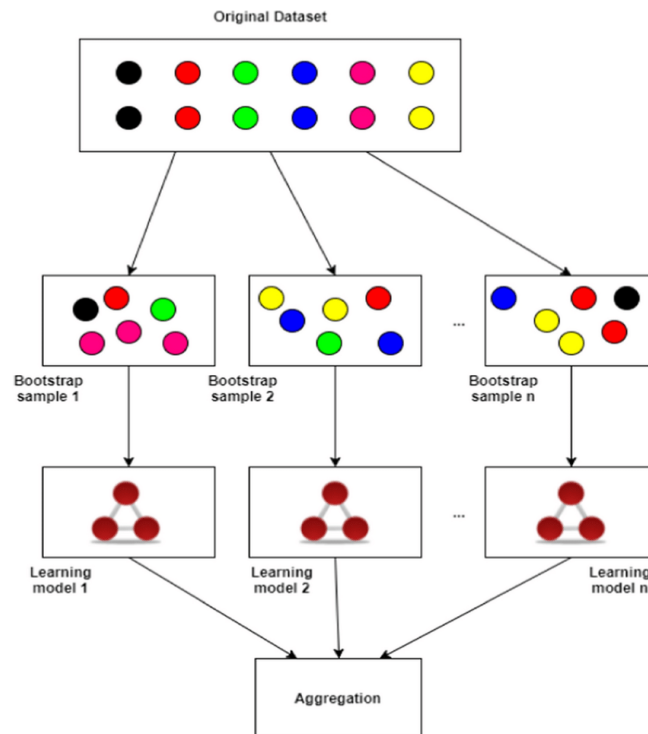
2.2.2 *Bootstrap Aggregating*

Metode lain yang juga dicetuskan oleh Leo Breiman adalah *bootstrap aggregating (bagging)* pada tahun 1996. Metode *bagging* menunjukkan dapat mengurangi variansi dari suatu estimator yang dapat mempertahankan kestabilan dan kekuatan dalam hasil prediksi (Fuleky, 2020). Mekanisme algoritma *bootstrap* bekerja dengan melakukan pengambilan sampel data dari variabel independen secara berulang-ulang dengan pengembalian, hal ini juga dikenal sebagai *resampling*. Sampel *bootstrap* banyak diambil dari data asli yang kemudian dilakukan evaluasi terhadap sampel-sampel tersebut guna mengetahui perkiraan standar *error* dari parameter yang diestimasi (Tibshirani & Efron, 1993). Setelah dilakukan *resampling*, kemudian dilakukan *aggregating*. *Aggregating* merupakan metode yang menggabungkan hasil prediksi dari masing-masing pohon klasifikasi yang telah terbentuk.

Proses dalam melakukan *bagging* adalah sebagai berikut:

1. Melakukan pengambilan sampel secara acak sebanyak n dengan pengembalian.
2. Melakukan penyusunan pohon berdasarkan hasil *resampling*.

3. Melakukan proses 1 – 2 hingga diperoleh jumlah pohon yang ditentukan dengan perulangan sebanyak n kali.
4. Melakukan penggabungan pohon pada proses 3 menggunakan *majority vote*.



Gambar 2.1 Proses *Random Forest*

2.2.3 Pembentukan Model

Model klasifikasi *random forest* umumnya dilakukan dari proses analisis *bootstrap aggregating (bagging)* yaitu ketika proses pemilihan pemilah. *Bootstrap aggregating* merupakan metode *ensemble* yang digunakan dengan tujuan untuk meningkatkan klasifikasi dengan melakukan kombinasi terhadap metode klasifikasi secara acak pada data training yang berfungsi untuk mengurangi variasi dan mencegah terjadinya *overfitting* (Nugroho & Religia, 2021). Algoritma pembentukan model *random forest* adalah sebagai berikut:

1. Melakukan pengambilan n data sampel dari data aktual dengan menggunakan *bootstrap resampling* dengan pengembalian. Proses perhitungan *bootstrap resampling*, yaitu:
 - a. Distribusi empiris \hat{F}_n dikonstruksi dari sampel acak dengan memberikan peluang $\frac{1}{n}$ pada setiap X_i dengan $i = 1, 2, \dots, n$.
 - b. *Bootstrap sample* dengan ukuran n diambil secara acak dengan pengembalian dari distribusi empiris \hat{F}_n yang kemudian disebut sebagai X^{*1} (*first bootstrap sample*).
 - c. Menghitung $\hat{\theta}_1^*$, yaitu statistik $\hat{\theta}$ yang diinginkan dari X^{*1} .
 - d. Melakukan pengulangan langkah b – c hingga Y kali hingga diperoleh $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_Y^*$.
 - e. Melakukan konstruksi terhadap distribusi peluang dari $\hat{\theta}_Y^*$ dengan memberikan peluang $\frac{1}{Y}$ pada setiap $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_Y^*$. Distribusi pada langkah ini merupakan estimator *bootstrap* untuk $\hat{\theta}$ yang dinotasikan sebagai \hat{F}^* .
 - f. Melakukan pendekatan terhadap estimasi *bootstrap*.

$$\hat{\theta}^* = \sum_{y=1}^Y \hat{\theta}_y^* \left(\frac{1}{Y} \right) \quad (2.13)$$

2. Melakukan penyusunan pohon klasifikasi dari hasil *bootstrap resampling* untuk setiap hasil dengan variabel prediktor terbaik sebagai penentu pemilah terbaik yang diambil secara *random* tanpa pemangkasan. Proses penyusunan pohon klasifikasi adalah sebagai berikut:
 - a. Menentukan *root node* dengan melakukan perhitungan terhadap nilai *entropy* dan *information gain*.

$$\text{Entropy}(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (2.14)$$

di mana:

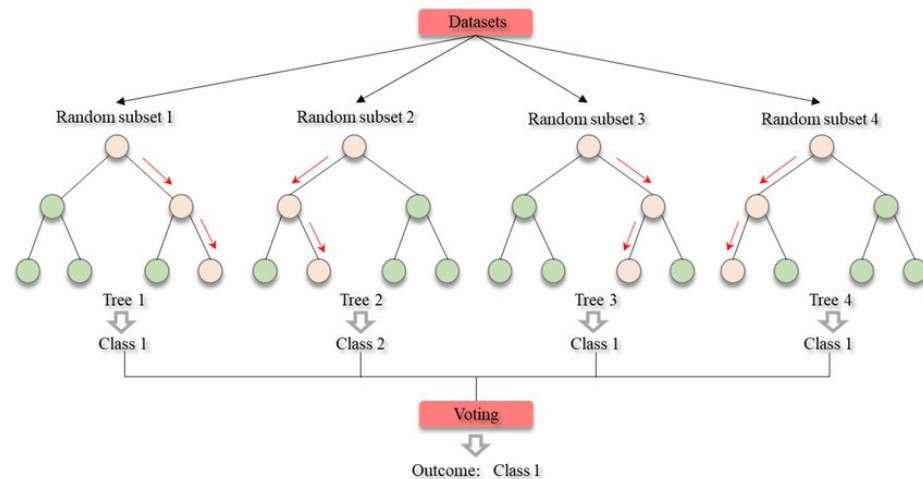
- S : Sampel Data yang Digunakan Saat *Training*
 p_i : Rasio Jumlah Sampel dari Subset dan Nilai Atribut ke- i
 c : Banyak Kelas

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (2.15)$$

di mana:

- S : Kelas dari Data
 A : Fitur
 $\text{Entropy}(S)$: *Entropy* dari Kelas S
 $\text{Values}(A)$: Banyaknya Atribut pada A
 S_v : Representasi Atribut dari A

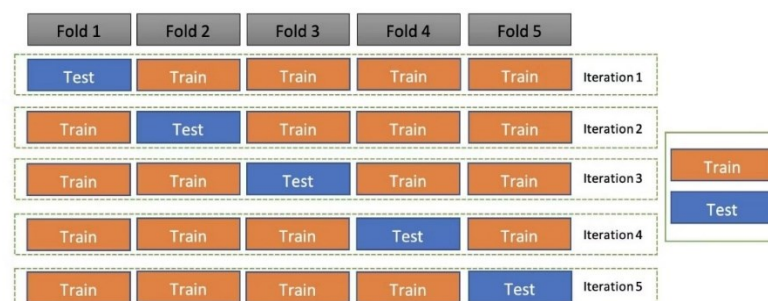
- b. Nilai *information gain* terbesar akan menjadi *root node*.
 - c. Melakukan proses a – b untuk memperoleh *branches*.
3. Melakukan prediksi model klasifikasi berdasarkan hasil dari pohon klasifikasi yang telah terbentuk.
 4. Melakukan proses 1 – 3 hingga diperoleh jumlah pohon klasifikasi yang diinginkan dengan perulangan yang dilakukan sebanyak n kali.
 5. Melakukan prediksi akhir klasifikasi dengan melakukan kombinasi terhadap hasil pohon klasifikasi yang telah diperoleh berdasarkan *majority vote*.



Gambar 2.2 Ilustrasi Pohon *Random Forest*

2.3 *K-Fold Cross Validation*

Metode *k-fold cross validation* merupakan metode untuk melakukan evaluasi terhadap model yang telah dibentuk untuk melihat kinerja yang dihasilkan. Metode ini dilakukan dengan melakukan pembagian data *training* dan data *testing* menjadi *k* (*fold*) bagian dengan besar yang sama. Kemudian model akan dilatih dengan menggunakan $k - 1$ *fold* dari data lalu diuji menggunakan data *testing* pada sisa satu bagian yang tidak digunakan saat pelatihan. Proses tersebut diulangi hingga sebanyak *k* kali dengan setiap bagian akan bergantian menjadi data *training*. Keuntungan menggunakan metode ini adalah bahwa setiap sampel pada data akan digunakan sebagai data *training* dan data *testing* setidaknya satu kali.



Gambar 2.3 Skema *K-Fold Cross Validation*

2.4 Pengukuran Tingkat Akurasi Klasifikasi

Menghitung ukuran ketepatan klasifikasi dalam analisis pengelompokan dilakukan dengan tujuan agar dapat mengetahui kebaikan model dalam klasifikasi data yang telah dilakukan. Terdapat banyak cara untuk menghitung ketepatan klasifikasi, salah satunya adalah dengan menggunakan *confusion matrix*. *Confusion matrix* adalah alat dalam analisis data klasifikasi yang digunakan untuk melakukan evaluasi hasil dari model klasifikasi yang telah dilakukan proses analisis (Khotimah & Utami, 2022). Konsep dasar dari *confusion matrix* adalah dengan melakukan perbandingan antara data aktual dengan data hasil analisis prediksi. Berikut merupakan tabel metode *confusion matrix* untuk permasalahan klasifikasi multikelas (Markoulidakis dkk., 2021).

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 hasil 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 (Hilmiyah, 2017).

Berdasarkan hasil dari *true positive* (TP), *false negative* (FN), *false positive* (FP), *true negative* (TN) dapat dicari nilai akurasi (*accuration*), sensitivitas (*recall*), presisi (*precission*) dan *f1 score*. *Accuration* menyatakan tingkat akurasi atau ketepatan hasil klasifikasi. *Precission* menunjukkan perbandingan antara kelas positif yang bernilai benar dengan seluruh kelas positif lainnya. Kemudian *recall* memperlihatkan presentase data dari kelas positif dengan hasil prediksi benar. *F₁Score* digunakan untuk melakukan evaluasi terhadap kinerja model klasifikasi (Sasongko, 2016). Persamaan untuk melakukan perhitungan *accuration*, *precision*, *sensitivity*, dan *f1 score* adalah sebagai berikut:

$$Accuracy = \frac{\sum_{i=1}^N TP(C_i)}{\sum_{i=1}^N \sum_{j=1}^N C_{i,j}} \times 100\% \quad (2.16)$$

$$Precision_{C_i} = \frac{TP(C_i)}{TP(C_i)+FP(C_i)} \times 100\% \quad (2.17)$$

$$Recall_{C_i} = \frac{TP(C_i)}{TP(C_i)+FN(C_i)} \times 100\% \quad (2.18)$$

$$F_1 \text{ Score} = 2 \times \left(\frac{precision \times recall}{precision + recall} \right) \quad (2.19)$$

2.5 Variable Importance

Variable importance merupakan ukuran yang menunjukkan pengaruh setiap variabel independen dalam proses pembuatan model klasifikasi. Tujuan dari *variable importance* adalah untuk memahami dan melakukan indentifikasi terhadap variabel yang memiliki pengaruh paling besar dalam proses klasifikasi. Salah satu cara untuk menentukan *variable importance* adalah dengan menggunakan *gini feature importance*. Keunggulan *gini importance* adalah sifatnya yang lebih sensitif dengan prediktor dengan satuan pengukuran yang berbeda sehingga dapat

memberikan informasi yang lebih akurat terhadap prediktor masing-masing variabel pada model klasifikasi (Nicodemus, 2011). Perhitungan *gini importance* dapat dilakukan dengan menggunakan formula sebagai berikut.

$$Imp(X_m) = \frac{1}{N_T} \sum_T \sum_{t \in T: v(s_t)=X_m} p(t) \Delta i(s_t, t) \quad (2.20)$$

di mana

- X_m : Variabel
- $p(t)$: Proporsi $\frac{N_T}{N}$ dari sampel yang mencapai t
- $p(t) \Delta i(s_t, t)$: Penurunan *weighted impurity*
- $v(s_t)$: Variabel yang digunakan dalam pemisahan s_t
- N_T : Pohon-pohon di hutan
- t : Node t

2.6 Ketimpangan Gender

Ketimpangan atau ketidakadilan gender merupakan fenomena di mana posisi perempuan memiliki tempat yang lebih rendah dibandingkan dengan laki-laki dalam banyak sektor di kehidupan sehari-hari (Apriliandra & Krisnani, 2020). Hal ini disebabkan karena masih banyak masyarakat yang memiliki persepsi bahwa perempuan memiliki kekuatan yang lebih kecil daripada laki-laki dalam berbagai aspek seperti pendidikan, pekerjaan, politik dan pemerintahan. Pemikiran seperti ini telah mengakar dalam pola pikir masyarakat dan menjadi sebuah bagian dari budaya yang percaya bahwa laki-laki memiliki kontrol penuh atas semuanya, hal ini disebut juga sebagai budaya patriarki (Irma & Hasanah, 2014).

Ketimpangan gender karena budaya patriarki berakibat pada perilaku diskriminatif yang dihadapi oleh perempuan. Salah satunya pada masalah sosial di

dalam masyarakat. Bentuk ketidakadilan gender pada aspek sosial adalah sebagai berikut:

1. Kekerasan dalam Rumah Tangga (KDRT)

Tahun 2016, Komnas Perempuan mempublikasikan terdapat sebanyak 245.548 kasus kekerasan terhadap perempuan yang tersebar di 34 provinsi di Indonesia. Perilaku ini disebabkan karena pemikiran konstruksi sosial bahwa laki-laki adalah pihak yang mendominasi dan perempuan adalah pihak yang lemah.

2. Pernikahan Dini

Menurut data penelitian yang dilakukan oleh Pusat Kajian Gender dan Seksualitas Universitas Indonesia pada tahun 2015, Indonesia merupakan negara dengan angka pernikahan dini peringkat kedua di wilayah Asia Tenggara. Konstruksi sosial dalam pernikahan, perempuan adalah pihak penerima nafkah dan hanya berperan dalam sektor domestik (rumah tangga). Pekerjaan utama perempuan adalah menjadi seorang istri dan ibu rumah tangga, tidak memiliki kesempatan untuk mengembangkan potensi pendidikan dan kemampuan yang mereka miliki.

3. Stigma Perceraian

Stigma di Negara Indonesia, ketika perceraian terjadi, pihak perempuan yang menjadi janda seringkali ditempatkan dalam posisi yang lemah, rendah, negatif, dan tidak berdaya. Posisi ini seringkali menimbulkan ketidakadilan dan diskriminasi terhadap perempuan.

4. Kasus Pelecehan Seksual

Pola pikir di mana ketika laki-laki dianggap berkaitan dengan ego maskulinitas dan feminitas dianggap sebagai sesuatu yang lemah. Ketika laki-laki dianggap sebagai pihak penggoda dan perempuan dianggap sebagai pihak yang memang pantas untuk digoda karena ia merupakan sebab dari tindakan itu sendiri.

Ketimpangan gender dapat terjadi disebabkan oleh beberapa faktor. Menurut (Hasibuan, 2006) perbedaan gender dapat menyebabkan ketimpangan gender, hal ini dapat dilihat dari berbagai manifestasinya, yaitu:

1. Marjinalisasi

Marjinalisasi merupakan suatu proses saat perempuan dipinggirkan sehingga terdapat batasan yang menyebabkan perempuan kesulitan untuk berekspresi, bertindak dan melakukan eksplorasi jati dirinya. Proses ini menempatkan kedudukan berdasarkan jenis kelamin, dan perempuan ditempatkan pada kedudukan yang tidak penting (Apriliandra & Krisnani, 2020).

2. Subordinasi

Subordinasi adalah perilaku diskriminatif dengan melakukan pengelompokan antara dua gender dengan menyematkan status sosial di salah satunya. Pada perempuan, hal ini dapat dikatakan sebagai bentuk penomorduaan karena perempuan diletakkan pada posisi yang lebih rendah dibandingkan dengan laki-laki. Contohnya adalah perempuan dianggap sebagai pribadi yang emosional dan tidak rasional, sehingga nilai perempuan jauh lebih rendah dari pada laki-laki untuk menjadi pemimpin.

3. *Stereotype*

Stereotype merupakan bentuk pelabelan yang umumnya bersifat negatif. Perempuan memiliki banyak *stereotype* yang mengakibatkan ketidakadilan dan diskriminasi yang merugikan perempuan. Contohnya adalah perempuan seharusnya memiliki sifat yang ramah, namun apabila perempuan ramah terhadap lawan jenis akan dianggap sebagai wanita yang merayu.

4. Kekerasan

Kekerasan adalah bentuk serangan yang dilakukan terhadap fisik maupun secara psikologis yang umumnya terjadi pada perempuan akibat perbedaan gender. Kekerasan terhadap perempuan ini disebabkan oleh *stereotype* yang dimiliki perempuan dalam masyarakat.

2.7 Indeks Khusus Ketimpangan Gender

Pemerintah Indonesia melakukan banyak upaya sebagai bentuk keseriusan dalam mengurangi ketimpangan gender sejak tahun 2000. Ketimpangan gender tidak hanya memengaruhi masalah sosial, namun juga berpengaruh pada sektor lainnya seperti pertumbuhan ekonomi, bidang kesehatan, bidang pendidikan dan ketenagakerjaan. Sebagai contoh dalam bidang kesehatan dan ekonomi, perempuan memiliki angka harapan hidup lebih tinggi daripada laki-laki, namun jika tidak disertai dengan optimalisasi peran perempuan dalam pembangunan, jumlah ketersediaan lapangan kerja untuk perempuan dan keahlian yang dimiliki perempuan dalam pekerjaan akan menurunkan tingkat produktivitas untuk menghasilkan jasa atau barang yang dapat menggerakkan roda ekonomi, sehingga

pertumbuhan perekonomian suatu negara tersebut akan terhambat (Deris dkk., 2022).

Pemerintah membentuk indeks ketimpangan gender (IKG) sebagai alat ukur untuk tingkat ketimpangan gender yang ada di Indonesia agar dapat dijadikan evaluasi terhadap program penanganan yang dilakukan (BPS, 2022). Indeks Ketimpangan Gender (IKG) adalah instrument khusus yang dibuat dengan tujuan untuk mengukur tingkat ketidaksetaraan gender yang terjadi di Indonesia. IKG disusun oleh dimensi yang terdiri atas beberapa indikator adalah sebagai berikut (BPS, 2022):

1. Dimensi kesehatan yang terdiri atas dua indikator, yaitu proporsi persalinan tidak dilakukan di fasilitas kesehatan dan proporsi wanita berusia 15-49 tahun yang pernah kawin dan saat melahirkan hidup pertama berusia kurang dari 20 tahun.
2. Dimensi pemberdayaan terdiri atas dua indikator, yaitu persentase penduduk laki-laki dan perempuan dengan jenjang pendidikan minimal SMA dan persentase laki-laki dan perempuan yang duduk di parlemen.
3. Dimensi pasar tenaga kerja terdiri atas satu indikator, yaitu tingkat partisipasi Angkatan kerja (TPAK) laki-laki dan perempuan.

Nilai IKG diperoleh dari beberapa perhitungan nilai indeks. Nilai indeks yang perlu dilakukan perhitungan adalah indeks perempuan, indeks laki-laki, agregasi indeks perempuan dan laki-laki, indeks kesehatan, indeks pemberdayaan, indeks pasar tenaga kerja, indeks seluruh dimensi, indeks kesetaraan gender dan indeks ketimpangan gender. Berikut persamaan untuk menghitung nilai dari masing-masing indeks (BPS, 2022).

Indeks perempuan:

$$G_F = \sqrt[3]{\left(\frac{0,01}{Faskes} \cdot \frac{0,01}{ULP}\right)^{\frac{1}{2}} \cdot (PR_F \cdot SE_F)^{\frac{1}{2}} \cdot TPAK_F} \quad (2.21)$$

Indeks laki-laki:

$$G_M = \sqrt[3]{1 \cdot (PR_M \cdot SE_M)^{\frac{1}{2}} \cdot TPAK_M} \quad (2.22)$$

Agregasi indeks perempuan dan laki-laki:

$$HARM(G_F, G_M) = \left[\frac{(G_F)^{-1} + (G_M)^{-1}}{2} \right]^{-1} \quad (2.23)$$

Indeks kesehatan:

$$Kesehatan = \frac{\left(\sqrt{\frac{0,01}{faskes} \cdot \frac{0,01}{ULP} + 1} \right)}{2} \quad (2.24)$$

Indeks pemberdayaan:

$$Pemberdayaan = \frac{\sqrt{PR_F \cdot SE_F} + \sqrt{PR_M \cdot SE_M}}{2} \quad (2.25)$$

Indeks pasar tenaga kerja:

$$IPK = \frac{TPAK_F + TPAK_M}{2} \quad (2.26)$$

Indeks seluruh dimensi:

$$G_{F,M} = \sqrt[3]{Kesehatan \cdot Pemberdayaan \cdot IPK} \quad (2.27)$$

Indeks kesetaraan gender:

$$Indeks Kesetaraan Gender = \frac{HARM(G_F, G_M)}{G_{F,M}} \quad (2.28)$$

Indeks ketimpangan gender:

$$IKG = 1 - \frac{HARM(G_F, G_M)}{G_{F,M}} \quad (2.29)$$

Keterangan:

G_F : Indeks Perempuan

G_M	: Indeks Laki-laki
Faskes	: Proporsi persalinan yang tidak di fasilitas kesehatan
ULP	: Proporsi 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 pendidikan minimal SMA
SE_M	: Persentase laki-laki 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
Kesehatan	: Indeks kesehatan reproduksi
Pemberdayaan	: Indeks pemberdayaan
IPK	: Indeks pasar kerja
$G_{F,M}$: Indeks dimensi

2.8 Kajian Integrasi Ketimpangan Gender dengan Al-Quran

Konsep gender dan jenis kelamin (*sex*) perlu dipahami dalam konteks ketimpangan gender. *Sex* merupakan alat yang digunakan sebagai pembeda antara pria dan wanita berdasarkan anatomi tubuh manusia yang bersifat tetap, kodrat dan tidak dapat ditukar (Subki dkk., 2021). Sedangkan *gender* adalah suatu konsep peran dan tanggung jawab pria dan wanita yang digunakan untuk menilai

kesenjangan antara laki-laki dan perempuan akibat dari rekonstruksi sosial budaya (Adam dkk., 2022).

Berkaitan dengan permasalahan antara hubungan laki-laki dan perempuan dalam hal kesetaraan gender, Al-Quran memaparkan sudut pandang secara *egaliter*. Sebagaimana diungkapkan dalam QS. Al-Maidah:08 pada arti berikut (Kemenag, 2024).

“Hai orang-orang yang beriman hendaklah kamu jadi orang-orang yang selalu menegakkan (kebenaran) karena Allah, menjadi saksi dengan adil. Dan janganlah sekali-kali kebencianmu terhadap sesuatu kaum, mendorong kamu untuk berlaku tidak adil. Berlaku adillah, karena adil itu lebih dekat kepada takwa. Dan bertakwalah kepada Allah, sesungguhnya Allah Maha Mengetahui apa yang kamu kerjakan.”

Selain QS. Al-Maidah:08, menurut tafsir Quraish Shihab terhadap QS. Al-Hujurat ayat 13, disebutkan bahwa ayat tersebut memiliki peralihan kepada penjelasan mengenai prinsip dasar dalam hubungan antar manusia karena menggunakan panggilan kepada manusia. Kalimat pertama pada ayat yang menyebutkan *sesungguhnya Kami telah menciptakan kamu dari seorang laki-laki dan perempuan* merupakan sebuah bentuk penegasan bahwa setiap manusia memiliki tingkat derajat yang sama dihadapan Allah SWT, tidak ada perbedaan antara perempuan dan laki-laki dalam hal nilai kemanusiaan. Kalimat terakhir pada ayat *sesungguhnya yang paling mulia diantara kamu di sisi Allah adalah orang yang paling bertaqwa* menunjukkan bahwa setiap manusia baik laki-laki maupun perempuan harus berusaha untuk meningkatkan ketaqwaan agar menjadi orang yang paling mulia disisi Allah (Subki dkk., 2021).

Berdasarkan tafsir oleh Quraish Shihab dapat dikatakan bahwa Islam sangat memuliakan baik laki-laki maupun perempuan. Tidak terdapat perbedaan antara laki-laki dan perempuan kecuali nilai ketaqwaannya kepada Allah SWT. Laki-laki

dan perempuan memiliki *level* yang sejajar dalam berbagai aspek kehidupan (Subki dkk., 2021). Sebagaimana dalam QS. Al-Hujurat ayat 13, Nabi Muhammad SAW memberikan hak yang setara antara pria dan wanita. Contohnya adalah baik laki-laki dan perempuan memiliki hak partisipasi yang sama dalam berperang, bermusyawarah, dan berpolitik. Ummu al-Mundzir binti Qays merupakan salah satu perempuan Anshar yang menyatakan baiat kepada Nabi Muhammad SAW dan ikut dalam perjuangan umat islam dalam mengusir kaum yahudi dan kaum kafir, kemudian Aisyah yang memiliki peran dalam kepemimpinan dalam perang jamal (Sholichah, 2021).

2.9 Kajian Ketimpangan Gender dengan Teori Pendukung

Penelitian ini dilakukan atas beberapa langkah penyelesaian. Pertama adalah melakukan pengumpulan data. Kedua yaitu melakukan analisis deskriptif yang bertujuan untuk mendapatkan representasi data secara umum terkait karakteristik data pada penelitian yakni Indeks Ketimpangan Gender pada 513 kabupaten/kota di Indonesia. Langkah ketiga yakni melakukan pemisahan data menjadi data *training* dan data *testing*. Kemudian membentuk model klasifikasi dengan metode *random forest* menggunakan data *training*. Tahap awal dalam membentuk model klasifikasi adalah dengan melakukan *bootstrap sampling* terhadap n data sampel *dataset* dengan pengembalian. Tahap kedua adalah menyusun pohon klasifikasi dari hasil *bagging* untuk setiap hasil sampel dan menentukan pemilah terbaik berdasarkan variabel prediktor. Kemudian menentukan prediksi terhadap model klasifikasi berdasarkan pohon klasifikasi. Model klasifikasi *random forest* yang telah terbentuk akan memberikan klasifikasi indeks ketimpangan gender pada 513 kabupaten/kota di Indonesia dalam empat kelas. Jika hasil *random forest* bernilai

rendah, maka akan termasuk dalam kategori rendah. Kemudian, jika hasil *random forest* bernilai menengah ke bawah, maka akan termasuk dalam kategori menengah ke bawah. Selanjutnya, apabila hasil *random forest* bernilai menengah ke atas, maka akan termasuk dalam kategori menengah ke atas. Terakhir, ketika hasil *random forest* bernilai tinggi, maka akan termasuk dalam kategori tinggi. Langkah selanjutnya setelah data selesai diklasifikasikan, dilakukan pengukuran ketepatan model *random forest* dengan menggunakan data *testing*. Pengukuran dilakukan dengan menggunakan *confusion matrix* untuk melihat tingkat akurasi yang dihasilkan saat analisis klasifikasi. Langkah terakhir dalam penelitian adalah melakukan analisis hasil prediksi model klasifikasi yang telah dihasilkan.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Jenis pendekatan yang digunakan pada penelitian ini adalah penelitian deskriptif kuantitatif. Penelitian ini menggunakan data kuantitatif yaitu data indeks ketimpangan gender pada 514 kabupaten/kota di Indonesia tahun 2021 yang selanjutnya dilakukan analisis menggunakan metode *random forest*.

3.2 Data dan Sumber Data

Penelitian ini menggunakan data berupa data sekunder. Data sekunder merupakan jenis data yang didapatkan melalui perantara dari sumber yang telah tersedia. Data penelitian dapat diakses melalui *website* resmi Badan Pusat Statistik (BPS) Indonesia.

3.3 Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder adalah jenis data yang diperoleh dari sumber yang sudah tersedia. Data penelitian dapat dilihat dan diakses melalui *website* Badan Pusat Statistik (BPS) Indonesia. Data penelitian yang digunakan terdiri atas variabel dependent (Y) dan variabel independent (X). Variabel yang akan digunakan dalam penelitian sebagai berikut:

Tabel 3.1 Variabel Penelitian

No.	Variabel		Skala	Keterangan
1.	Indeks Ketimpangan Gender (IKG) pada 513 Kabupaten/Kota di Indonesia	(Y)	Ordinal	Rendah: (IKG < 0,399) Menengah Bawah: (0,450 < IKG < 0,499) Menengah Atas: (0,450 < IKG < 0,500) Tinggi: (IKG ≥ 0,500) Sumber: BPS, 2022
2.	Persalinan Tidak di Fasilitas Kesehatan	(X ₁)	Rasio	0 – 100
3.	Wanita Berusia 15 – 49 Tahun yang Pernah Kawin dan Saat Melahirkan Hidup Pertama Berusia Kurang dari 20 Tahun	(X ₂)	Rasio	0 – 100
4.	Laki-Laki yang Duduk di Parlemen	(X ₃)	Rasio	0 – 100
5.	Perempuan yang Duduk di Parlemen	(X ₄)	Rasio	0 – 100
6.	Penduduk Laki-Laki dengan Pendidikan Minimal SMA	(X ₅)	Rasio	0 – 100
7.	Penduduk Perempuan dengan Pendidikan Minimal SMA	(X ₆)	Rasio	0 – 100
8.	Angkatan Kerja Terhadap Penduduk Usia Kerja (Laki-laki)	(X ₇)	Rasio	0 – 100
9.	Angkatan Kerja Terhadap Penduduk Usia Kerja (Perempuan)	(X ₈)	Rasio	0 – 100

3.4 Tahapan Penelitian

Adapun langkah-langkah yang dilakukan dalam penelitian adalah sebagai berikut:

1. Melakukan pengumpulan terhadap data klasifikasi ketimpangan gender yang terdiri atas variabel *dependent* dan variabel *independent*.

2. Melakukan analisis deskriptif pada data indeks ketimpangan gender.
 3. Melakukan pembagian data yakni 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. Melakukan *balancing data* terhadap data *training* dengan menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE).
 5. Membentuk model *random forest* dengan menggunakan data *training*.
- Berikut merupakan perhitungan secara manual proses klasifikasi menggunakan metode *random forest* dengan menggunakan 8 sampel data.
- a. Melakukan *bootstrap* sampling untuk mengambil n data sampel dengan pengembalian.

Tabel 3.2 Contoh Data Perhitungan Manual

Wilayah	Kelas	X1	X2	X5	X6
Kota Jakarta Pusat	Rendah	0,10	10,20	69,28	63,20
Ngawi	Rendah	0,10	20,70	31,00	28,13
Bandung	Menengah Bawah	9,70	30,80	41,80	35,74
Aceh Besar	Menengah Bawah	4,40	18,50	53,69	52,86
Pamekasan	Menengah Atas	8,30	32,90	35,99	21,87
Tapanuli Selatan	Menengah Atas	46,00	15,20	38,82	33,19
Raja Ampat	Tinggi	77,30	31,00	39,07	32,08
Bogor	Tinggi	25,50	32,80	39,94	30,62

Proses *bootstrap sampling* dilakukan setiap suatu pohon keputusan akan dibangun. Berikut merupakan data *training* dari data sampel.

Tabel 3.3 Data Training Perhitungan Manual

Wilayah	Kelas	X1	X2	X5	X6
Tapanuli Selatan	Menengah Atas	46	15,2	38,82	33,19
Bandung	Menengah Bawah	9,7	30,8	41,8	35,74
Ngawi	Rendah	0,1	20,7	31	28,13
Aceh Besar	Menengah Bawah	4,4	18,5	53,69	52,86
Pamekasan	Menengah Atas	8,3	32,9	35,99	21,87
Bogor	Tinggi	25,5	32,8	39,94	30,62

Proses *bootstrapping sampling* dilakukan menggunakan data *training* dengan pengambilan sampel yang dilakukan secara acak dengan pengembalian. Proses *bootstrapping sampling* akan selalu dilakukan ketika akan membangun sebuah pohon keputusan (*decision tree*) sehingga pada setiap pohon nantinya akan memiliki sampel yang berbeda.

Tabel 3.4 *Bootstrap* Pertama

Wilayah	Kelas	X1	X2	X5	X6
Pamekasan	Menengah Atas	8,3	32,9	35,99	21,87
Bogor	Tinggi	25,5	32,8	39,94	30,62
Tapanuli Selatan	Menengah Atas	46	15,2	38,82	33,19
Bandung	Menengah Bawah	9,7	30,8	41,8	35,74
Bandung	Menengah Bawah	9,7	30,8	41,8	35,74
Aceh Besar	Menengah Bawah	4,4	18,5	53,69	52,86

Berdasarkan Tabel 3.4, terdapat tiga kabupaten/kota dengan kelas menengah bawah, dua kabupaten/kota dengan kelas menengah atas, dan satu kabupaten/kota dengan kelas tinggi.

- b. Menyusun pohon klasifikasi dari hasil *bootstrap sampling* dengan penentu pemilah adalah variabel prediktor terbaik yang didapatkan dari nilai *entropy* dan *information gain* serta tanpa dilakukan pemangkasan.

Setelah data *bootstrap sampling* terbentuk, langkah berikutnya adalah membentuk pohon klasifikasi dengan menggunakan variabel

prediktor terbaik sebagai penentu pemilah pada setiap node serta dilakukan tanpa pemangkasan pada pohon. *Root node* pada pohon klasifikasi diperoleh dengan melakukan perhitungan terhadap *entropy value* dan *information gain* setiap variabel pada data *bootstrap sampling*. Berikut adalah perhitungan manual untuk menemukan nilai *entropy* semua data terhadap komposisi kelas pada data *bootstrap* pertama.

$$\begin{aligned} \text{Entropy}(S) &= -\left((p(\text{menengah bawah}|\text{semua}) \times \right. \\ &\quad \left. \log_2 p(\text{menengah bawah}|\text{semua})) + \right. \\ &\quad \left. (p(\text{menengah atas}|\text{semua}) \times \right. \\ &\quad \left. \log_2 p(\text{menengah atas}|\text{semua})) + \right. \\ &\quad \left. (p(\text{tinggi}|\text{semua}) \times \log_2 p(\text{tinggi}|\text{semua})) \right) \end{aligned}$$

$$\begin{aligned} \text{Entropy}(S) &= -\left(\left(\frac{2}{6} \times \log_2 \frac{2}{6} \right) + \left(\frac{3}{6} \times \log_2 \frac{3}{6} \right) + \left(\frac{1}{6} \times \log_2 \frac{1}{6} \right) \right) \\ &= -\left(\left(\frac{3}{6} \times \log_2 \frac{3}{6} \right) + \left(\frac{2}{6} \times \log_2 \frac{2}{6} \right) + \left(\frac{1}{6} \times \log_2 \frac{1}{6} \right) \right) \\ &= -\left((0,5 \times \log_2 0,5) + (0,333 \times \log_2 0,333) + \right. \\ &\quad \left. (0,1667 \times \log_2 0,1667) \right) \\ &= -\left((0,5 \times (-1)) + (0,333 \times (-1,585)) + \right. \\ &\quad \left. (0,1667 \times (-2,585)) \right) \\ &= -\left((-0,5) + (-0,5278) + (-0,431) \right) \\ &= -(-1,4588) \end{aligned}$$

$$\text{Entropy}(S) = 1,4588$$

Setelah nilai *entropy* semua data terhadap komposisi kelas pada data *bootstrap* pertama ditemukan, kemudian melakukan pembentukan

node pertama dari pohon. Pada *node* pertama ini akan dilakukan perhitungan dari variabel X_1 dan X_2 . Sebelum perhitungan *entropy*, perlu dilakukan proses diskritisasi pada variabel agar memudahkan proses *splitting* guna menemukan variabel prediktor untuk *node* pertama. Proses diskritisasi menggunakan metode *binning*, yaitu mengelompokkan data kontinu dalam “*bins*” yang lebih kecil. Proses dalam melakukan metode *binning* adalah sebagai berikut:

- Menentukan jumlah *bin* yang akan digunakan.
- Mengurutkan data dari angka yang terkecil hingga yang terbesar.
- Melakukan perhitungan terhadap nilai rata-rata masing-masing *bin*.

Berikut merupakan proses perhitungan manual untuk menemukan nilai *entropy* dan *information gain*, sekaligus dengan proses diskritisasi di dalamnya.

Tabel 3.5 *Bins* Variabel X_1

Wilayah	Kelas	X1	Mean	Jumlah ≤	Jumlah >
Aceh Besar	Menengah Bawah	4,4	6,35	1	5
Pamekasan	Menengah Atas	8,3			
Bandung	Menengah Bawah	9,7	9,7	4	2
Bandung	Menengah Bawah	9,7			
Bogor	Tinggi	25,5	25,5	5	1
Tapanuli Selatan	Menengah Atas	46			

- Perhitungan *Entropy* dan *Information Gain* Variabel X_1

$$\begin{aligned}
 E(\leq 6,35) &= -((p(\text{menengah bawah}|semua) \times \\
 &\quad \log_2 p(\text{menengah bawah}|semua))) \\
 &= -\left(\frac{1}{1} \times \log_2 \frac{1}{1}\right)
 \end{aligned}$$

$$E(\leq 6,35) = -(1 \times 0) \\ = 0$$

$$E(> 6,35) = -((p(\text{menengah bawah}|semua) \times \\ \log_2 p(\text{menengah bawah}|semua)) + \\ (p(\text{menengah atas}|semua) \times \\ \log_2 p(\text{menengah atas}|semua))) \\ = -\left(\left(\frac{2}{5} \times \log_2 \frac{2}{5}\right) + \left(\frac{2}{5} \times \log_2 \frac{2}{5}\right) + \right. \\ \left. \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right)\right) \\ = -((0,40 \times \log_2 0,40) + (0,40 \times \\ \log_2 0,40) + (0,20 \times \log_2 0,20)) \\ = -((0,40 \times (-1,322)) + (0,40 \times \\ (-1,322)) + (0,20 \times (-2,322))) \\ = -((-0,5288) + (-0,5288) + (-0,4644)) \\ = -(-1,522) \\ = 1,522$$

$$E(\leq 9,7) = -((p(\text{menengah bawah}|semua) \times \\ \log_2 p(\text{menengah bawah}|semua)) + \\ (p(\text{menengah atas}|semua) \times \\ \log_2 p(\text{menengah atas}|semua)) + \\ (p(\text{tinggi}|semua) \times \\ \log_2 p(\text{tinggi}|semua))) \\ = -\left(\left(\frac{3}{4} \times \log_2 \frac{3}{4}\right) + \left(\frac{1}{4} \times \log_2 \frac{1}{4}\right)\right)$$

$$\begin{aligned}
E(\leq 9,7) &= -((0,75 \times \log_2 0,75) + (0,25 \times \log_2 0,25)) \\
&= -((0,75 \times (-0,415)) + (0,25 \times (-2))) \\
&= -((-0,311) + (-0,5)) \\
&= -(-0,811) \\
&= 0,811
\end{aligned}$$

$$\begin{aligned}
E(> 9,7) &= -(p(\text{menengah atas}|\text{semua}) \times \\
&\quad \log_2 p(\text{menengah atas}|\text{semua})) + \\
&\quad (p(\text{tinggi}|\text{semua}) \times \\
&\quad \log_2 p(\text{tinggi}|\text{semua})) \\
&= -\left(\left(\frac{1}{2} \times \log_2 \frac{1}{2}\right) + \left(\frac{1}{2} \times \log_2 \frac{1}{2}\right)\right) \\
&= -((0,50 \times \log_2 0,50) + (0,50 \times \log_2 0,50)) \\
&= -((0,50 \times (-1)) + (0,50 \times (-1))) \\
&= -(-0,50 + (-0,50)) \\
&= 1
\end{aligned}$$

$$\begin{aligned}
E(\leq 25.5) &= -(p(\text{menengah bawah}|\text{semua}) \times \\
&\quad \log_2 p(\text{menengah bawah}|\text{semua})) + \\
&\quad (p(\text{menengah atas}|\text{semua}) \times \\
&\quad \log_2 p(\text{menengah atas}|\text{semua})) + \\
&\quad (p(\text{tinggi}|\text{semua}) \times \\
&\quad \log_2 p(\text{tinggi}|\text{semua})) \\
&= -\left(\left(\frac{3}{5} \times \log_2 \frac{3}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right) + \right. \\
&\quad \left. \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right)\right)
\end{aligned}$$

$$\begin{aligned}
E(\leq 25,5) &= -((0,60 \times \log_2 0,60) + (0,20 \times \\
&\quad \log_2 0,20) + (0,20 \times \log_2 0,20)) \\
&= -((0,60 \times (-0,737)) + (0,20 \times \\
&\quad (-2,322)) + (0,20 \times (-2,322))) \\
&= -(-0,4422 + (-0,4644) + (-0,4644)) \\
&= -(-1,371) \\
&= 1,371
\end{aligned}$$

$$\begin{aligned}
E(> 25,5) &= -(p(\text{menengah atas}|\text{semua}) \times \\
&\quad \log_2 p(\text{menengah atas}|\text{semua})) \\
&= -\left(\frac{1}{1} \times \log_2 \frac{1}{1}\right) \\
&= -(1 \times 0) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\text{Gain}(\text{semua}, 63,5) &= 1,4588 - \left(\left(\frac{1}{6}\right) \times 0\right) + \\
&\quad \left(\left(\frac{5}{6}\right) \times 1,522\right) \\
&= 1,4588 - 0 + 1,2678 \\
&= 2,7266
\end{aligned}$$

$$\begin{aligned}
\text{Gain}(\text{semua}, 9,7) &= 1,4588 - \left(\left(\frac{4}{6}\right) \times 0,811\right) + \\
&\quad \left(\left(\frac{2}{6}\right) \times 1\right) \\
&= 1,4588 - 0,5407 + 0,3333 \\
&= 1,2514
\end{aligned}$$

$$\begin{aligned}
 \text{Gain}(\text{semua}, 25,5) &= 1,4588 - \left(\left(\frac{5}{6} \right) \times 1,371 \right) + \\
 &\quad \left(\left(\frac{1}{6} \right) \times 0 \right) \\
 &= 1,4588 - 1,1424 + 0 \\
 &= 0,3164
 \end{aligned}$$

Tabel 3.6 Entropy dan Gain Value X_1

Mean	Jumlah Kasus		Entropy	Gain
6,35	\leq	1	0	2,7266
	$>$	5	1,522	
9,7	\leq	4	0,811	1,2514
	$>$	2	1	
25,5	\leq	5	1,371	0,3164
	$>$	1	0	

b) Perhitungan Entropy dan Information Gain Variabel X_2

Tabel 3.7 Bins Variabel X_2

Wilayah	Kelas	X2	Mean	Jumlah	Jumlah
				\leq	$>$
Tapanuli Selatan	Menengah Atas	15,2	16,85	1	5
Aceh Besar	Menengah Bawah	18,5			
Bandung	Menengah Bawah	30,8	30,8	4	2
Bandung	Menengah Bawah	30,8			
Bogor	Tinggi	32,8	32,85	5	1
Pamekasan	Menengah Atas	32,9			

$$\begin{aligned}
 E(\leq 16,85) &= -((p(\text{menengah atas}|\text{semua}) \times \\
 &\quad \log_2 p(\text{menengah atas}|\text{semua})) \\
 &= -\left(\frac{1}{1} \times \log_2 \frac{1}{1} \right) \\
 &= -(1 \times 0) \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
E(> 16,85) &= -\left((p(\text{menengah bawah}|semua) \times \right. \\
&\quad \left. \log_2 p(\text{menengah bawah}|semua)) + \right. \\
&\quad \left. (p(\text{menengah atas}|semua) \times \right. \\
&\quad \left. \log_2 p(\text{menengah atas}|semua)) + \right. \\
&\quad \left. (p(\text{tinggi}|semua) \times \right. \\
&\quad \left. \log_2 p(\text{tinggi}|semua))\right)
\end{aligned}$$

$$\begin{aligned}
E(> 16,85) &= -\left(\left(\frac{3}{5} \times \log_2 \frac{3}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right) + \right. \\
&\quad \left. \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right)\right) \\
&= -\left((0,60 \times \log_2 0,60) + (0,20 \times \right. \\
&\quad \left. \log_2 0,20) + (0,20 \times \log_2 0,20)\right) \\
&= -\left((0,60 \times (-0,737)) + (0,20 \times \right. \\
&\quad \left. (-2,322)) + (0,20 \times (-2,322))\right) \\
&= -(-0,4422 + (-0,4644) + (-0,4644)) \\
&= -(-1,371) \\
&= 1,371
\end{aligned}$$

$$\begin{aligned}
E(\leq 30,8) &= -\left((p(\text{menengah bawah}|semua) \times \right. \\
&\quad \left. \log_2 p(\text{menengah bawah}|semua)) + \right. \\
&\quad \left. (p(\text{menengah atas}|semua) \times \right. \\
&\quad \left. \log_2 p(\text{menengah atas}|semua))\right) \\
&= -\left(\left(\frac{3}{4} \times \log_2 \frac{3}{4}\right) + \left(\frac{1}{4} \times \log_2 \frac{1}{4}\right)\right)
\end{aligned}$$

$$\begin{aligned}
E(\leq 30,8) &= -((0,75 \times \log_2 0,75) + (0,25 \times \\
&\quad \log_2 0,25)) \\
&= -((0,75 \times (-0,415)) + (0,25 \times (-2))) \\
&= -((-0,311) + (-0,5)) \\
&= -(-0,811) \\
&= 0,811
\end{aligned}$$

$$\begin{aligned}
E(> 30,8) &= -(p(\text{menengah atas}|\text{semua}) \times \\
&\quad \log_2 p(\text{menengah atas}|\text{semua})) + \\
&\quad (p(\text{tinggi}|\text{semua}) \times \\
&\quad \log_2 p(\text{tinggi}|\text{semua})) \\
&= -\left(\left(\frac{1}{2} \times \log_2 \frac{1}{2}\right) + \left(\frac{1}{2} \times \log_2 \frac{1}{2}\right)\right) \\
&= -((0,50 \times \log_2 0,50) + (0,50 \times \\
&\quad \log_2 0,50)) \\
&= -((0,50 \times (-1)) + (0,50 \times (-1))) \\
&= -(-0,50 + (-0,50)) \\
&= -(-1) \\
&= 1
\end{aligned}$$

$$\begin{aligned}
E(\leq 32,85) &= -(p(\text{menengah bawah}|\text{semua}) \times \\
&\quad \log_2 p(\text{menengah bawah}|\text{semua})) + \\
&\quad (p(\text{menengah atas}|\text{semua}) \times \\
&\quad \log_2 p(\text{menengah atas}|\text{semua})) +
\end{aligned}$$

$$(p(\text{tinggi}|\text{semua}) \times \\ \log_2 p(\text{tinggi}|\text{semua}))$$

$$E(\leq 32,85) = -\left(\left(\frac{3}{5} \times \log_2 \frac{3}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right)\right)$$

$$= -\left((0,60 \times \log_2 0,60) + (0,20 \times \log_2 0,20) + (0,20 \times \log_2 0,20)\right)$$

$$E(\leq 32,85) = -\left((0,60 \times (-0,737)) + (0,20 \times (-2,322)) + (0,20 \times (-2,322))\right)$$

$$= -(-0,4422 + (-0,4644) + (-0,4644))$$

$$= -(-1,371)$$

$$= 1,371$$

$$E(> 32,85) = -\left((p(\text{menengah atas}|\text{semua}) \times \log_2 p(\text{menengah atas}|\text{semua}))\right)$$

$$= -\left(\frac{1}{1} \times \log_2 \frac{1}{1}\right)$$

$$= -(1 \times 0)$$

$$E(> 32,85) = 0$$

$$\text{Gain}(\text{semua}, 16,85) = 1,4588 - \left(\left(\frac{1}{6}\right) \times 0\right) + \left(\left(\frac{5}{6}\right) \times 1,371\right)$$

$$= 1,4588 - 0 + 1,1424$$

$$= 0,3164$$

$$\begin{aligned}
 \text{Gain}(\text{semua}, 30,8) &= 1,4588 - \left(\left(\frac{4}{6} \right) \times 0,811 \right) + \\
 &\quad \left(\left(\frac{2}{6} \right) \times 1 \right) \\
 &= 1,4588 - 0,5407 + 0,3333 \\
 &= 1,2514
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(\text{semua}, 32,85) &= 1,4588 - \left(\left(\frac{5}{6} \right) \times 1,371 \right) + \\
 &\quad \left(\left(\frac{1}{6} \right) \times 0 \right) \\
 &= 1,4588 - 1,1424 + 0
 \end{aligned}$$

$$(\text{semua}, 32,85) = 0,3164$$

Tabel 3.8 Entropy dan Gain Value X_2

Mean		Jumlah Kasus	Entropy	Gain
16,85	\leq	1	0	0,3164
	$>$	5	1,371	
30,8	\leq	4	0,811	1,2514
	$>$	2	1	
32,85	\leq	5	1,371	0,3164
	$>$	1	0	

Berdasarkan perhitungan di atas, nilai *information gain* tertinggi terdapat pada variabel X_1 yaitu persalinan tidak di fasilitas kesehatan dengan nilai sebesar 2,7266. Dengan demikian, maka X_1 menjadi *node* pohon pertama. Setelah itu, akan dilakukan perhitungan untuk mencari cabang pohon dengan menggunakan variabel X_5 dan X_6 . Berikut merupakan perhitungan manual nilai *entropy* dan *information gain* untuk variabel X_5 dan X_6 .

c) Perhitungan Entropy dan Information Gain Variabel X_5

Tabel 3.9 Bins Variabel X_5

Wilayah	Kelas	X5	Mean	Jumlah \leq	Jumlah $>$
Pamekasan	Menengah Atas	35,99	37,40	1	5
Tapanuli Selatan	Menengah Atas	38,82			
Bogor	Tinggi	39,94	40,87	3	3
Bandung	Menengah Bawah	41,8			
Bandung	Menengah Bawah	41,8	47,75	5	1
Aceh Besar	Menengah Bawah	53,69			

$$\begin{aligned}
 E(\leq 37,40) &= -((p(\text{menengah atas}|\text{semua}) \times \\
 &\quad \log_2 p(\text{menengah atas}|\text{semua})) \\
 &= -\left(\frac{1}{1} \times \log_2 \frac{1}{1}\right) \\
 &= -(1 \times 0) \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 E(> 37,40) &= -((p(\text{menengah bawah}|\text{semua}) \times \\
 &\quad \log_2 p(\text{menengah bawah}|\text{semua})) + \\
 &\quad (p(\text{menengah atas}|\text{semua}) \times \\
 &\quad \log_2 p(\text{menengah atas}|\text{semua})) + \\
 &\quad (p(\text{tinggi}|\text{semua}) \times \log_2 p(\text{tinggi}|\text{semua}))) \\
 &= -\left(\left(\frac{3}{5} \times \log_2 \frac{3}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right) + \right. \\
 &\quad \left. \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right)\right)
 \end{aligned}$$

$$\begin{aligned}
E(> 37,40) &= -((0,60 \times \log_2 0,60) + (0,20 \times \log_2 0,20) + \\
&\quad (0,20 \times \log_2 0,20)) \\
&= -((0,60 \times (-0,737)) + (0,20 \times (-2,322)) + \\
&\quad (0,20 \times (-2,322))) \\
&= -(-0,4422 + (-0,4644) + (-0,4644)) \\
&= -(-1,371) \\
&= 1,371
\end{aligned}$$

$$\begin{aligned}
E(\leq 40,87) &= -((p(\text{menengah atas}|semua) \times \\
&\quad \log_2 p(\text{menengah atas}|semua)) + \\
&\quad (p(\text{tinggi}|semua) \times \\
&\quad \log_2 p(\text{tinggi}|semua))) \\
&= -\left(\left(\frac{2}{3} \times \log_2 \frac{2}{3}\right) + \left(\frac{1}{3} \times \log_2 \frac{1}{3}\right)\right) \\
&= -((0,67 \times \log_2 0,67) + (0,33 \times \\
&\quad \log_2 0,33)) \\
&= -((0,67 \times (-0,577)) + (0,33 \times (-1,599))) \\
&= -((-0,386) + (-0,527)) \\
&= -(-0,913) \\
&= 0,913
\end{aligned}$$

$$\begin{aligned}
E(> 40,87) &= -((p(\text{menengah bawah}|semua) \times \\
&\quad \log_2 p(\text{menengah bawah}|semua))) \\
&= -\left(\frac{3}{3} \times \log_2 \frac{3}{3}\right) \\
&= -(1 \times \log_2 1)
\end{aligned}$$

$$E(> 40,87) = -(1 \times 0)$$

$$= 0$$

$$E(\leq 47,75) = -((p(\text{menengah bawah}|semua) \times \log_2 p(\text{menengah bawah}|semua)) + (p(\text{menengah atas}|semua) \times \log_2 p(\text{menengah atas}|semua)) + (p(\text{tinggi}|semua) \times \log_2 p(\text{tinggi}|semua)))$$

$$= -\left(\left(\frac{2}{5} \times \log_2 \frac{2}{5}\right) + \left(\frac{2}{5} \times \log_2 \frac{2}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right)\right)$$

$$= -((0,40 \times \log_2 0,40) + (0,40 \times \log_2 0,40) + (0,20 \times \log_2 0,20))$$

$$= ((0,40 \times (-1,322)) + (0,40 \times (-1,322)) + (0,20 \times (-2,322)))$$

$$= -((-0,5288) + (-0,5288) + (-0,4644))$$

$$= -(-1,522)$$

$$= 1,522$$

$$E(> 47,75) = -(p(\text{menengah bawah}|semua) \times \log_2 p(\text{menengah bawah}|semua))$$

$$= -\left(\frac{1}{1} \times \log_2 \frac{1}{1}\right)$$

$$= -(1 \times 0)$$

$$= 0$$

$$\begin{aligned}
 \text{Gain}(\text{semua}, 37,40) &= 1,4588 - \left(\left(\frac{1}{6} \right) \times 0 \right) + \\
 &\quad \left(\left(\frac{5}{6} \right) \times 1,371 \right) \\
 &= 1,4588 - 0 + 1,1425 \\
 &= 2,6013
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(\text{semua}, 40,87) &= 1,4588 - \left(\left(\frac{3}{6} \right) \times 0,913 \right) + \\
 &\quad \left(\left(\frac{3}{6} \right) \times 0 \right)
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(\text{semua}, 40,87) &= 1,4588 - 0,4565 + 0 \\
 &= 1,0023
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(\text{semua}, 47,75) &= 1,4588 - \left(\left(\frac{5}{6} \right) \times 1,522 \right) + \\
 &\quad \left(\left(\frac{1}{6} \right) \times 0 \right) \\
 &= 1,4588 - 1,268 + 0
 \end{aligned}$$

$$\text{Gain}(\text{semua}, 47,75) = 0,1908$$

Tabel 3.10 Entropy dan Gain Value X_5

Mean	Jumlah Kasus	Entropy	Gain
37,40	≤	1	2,6013
	>	5	
40,87	≤	3	1,0023
	>	3	
47,75	≤	5	0,1908
	>	1	

d) Perhitungan Entropy dan Information Gain Variabel X_6

Tabel 3.11 Bins Variabel X_6

Wilayah	Kelas	X2	Mean	Jumlah \leq	Jumlah $>$
Tapanuli Selatan	Menengah Atas	15,2	16,85	1	5
Aceh Besar	Menengah Bawah	18,5			
Bandung	Menengah Bawah	30,8	30,8	4	2
Bandung	Menengah Bawah	30,8			
Bogor	Tinggi	32,8	32,85	5	1
Pamekasan	Menengah Atas	32,9			

$$\begin{aligned}
 E(\leq 16,85) &= -((p(\text{menengah atas}|\text{semua}) \times \\
 &\quad \log_2 p(\text{menengah atas}|\text{semua})) \\
 &= -\left(\frac{1}{1} \times \log_2 \frac{1}{1}\right) \\
 &= -(1 \times 0) \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 E(> 16,85) &= -((p(\text{menengah bawah}|\text{semua}) \times \\
 &\quad \log_2 p(\text{menengah bawah}|\text{semua})) + \\
 &\quad (p(\text{menengah atas}|\text{semua}) \times \\
 &\quad \log_2 p(\text{menengah atas}|\text{semua})) + \\
 &\quad (p(\text{tinggi}|\text{semua}) \times \log_2 p(\text{tinggi}|\text{semua}))) \\
 &= -\left(\left(\frac{3}{5} \times \log_2 \frac{3}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right)\right) \\
 &= -((0,60 \times \log_2 0,60) + (0,20 \times \log_2 0,20) + \\
 &\quad (0,20 \times \log_2 0,20))
 \end{aligned}$$

$$\begin{aligned}
E(> 16,85) &= -\left((0,60 \times (-0,737)) + (0,20 \times (-2,322)) +\right. \\
&\quad \left.(0,20 \times (-2,322))\right) \\
&= -(-0,4422 + (-0,4644) + (-0,4644)) \\
&= -(-1,371) \\
&= 1,371
\end{aligned}$$

$$\begin{aligned}
E(\leq 30,8) &= -\left(p(\text{menengah bawah}|semua) \times\right. \\
&\quad \log_2 p(\text{menengah bawah}|semua)) + \\
&\quad \left(p(\text{menengah atas}|semua) \times\right. \\
&\quad \left.\log_2 p(\text{menengah atas}|semua))\right) \\
&= -\left(\left(\frac{3}{4} \times \log_2 \frac{3}{4}\right) + \left(\frac{1}{4} \times \log_2 \frac{1}{4}\right)\right) \\
&= -\left((0,75 \times \log_2 0,75) + (0,25 \times \log_2 0,25)\right) \\
&= -\left((0,75 \times (-0,415)) + (0,25 \times (-2))\right) \\
&= -(-0,311 + (-0,5)) \\
&= -(-0,811) \\
&= 0,811
\end{aligned}$$

$$\begin{aligned}
E(> 30,8) &= -\left(p(\text{menengah atas}|semua) \times\right. \\
&\quad \log_2 p(\text{menengah atas}|semua)) + \\
&\quad \left(p(\text{tinggi}|semua) \times \log_2 p(\text{tinggi}|semua))\right) \\
&= -\left(\left(\frac{1}{2} \times \log_2 \frac{1}{2}\right) + \left(\frac{1}{2} \times \log_2 \frac{1}{2}\right)\right) \\
&= -\left((0,50 \times \log_2 0,50) + (0,50 \times \log_2 0,50)\right) \\
&= -\left((0,50 \times (-1)) + (0,50 \times (-1))\right)
\end{aligned}$$

$$\begin{aligned}
E(> 30,8) &= -(-0,50 + (-0,50)) \\
&= -(-1) \\
&= 1
\end{aligned}$$

$$\begin{aligned}
E(\leq 32,85) &= -(p(\text{menengah bawah}|semua) \times \\
&\quad \log_2 p(\text{menengah bawah}|semua)) + \\
&\quad (p(\text{menengah atas}|semua) \times \\
&\quad \log_2 p(\text{menengah atas}|semua)) + \\
&\quad (p(\text{tinggi}|semua) \times \log_2 p(\text{tinggi}|semua))) \\
&= -\left(\left(\frac{3}{5} \times \log_2 \frac{3}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right)\right) \\
&= -((0,60 \times \log_2 0,60) + (0,20 \times \log_2 0,20) + \\
&\quad (0,20 \times \log_2 0,20)) \\
&= -((0,60 \times (-0,737)) + (0,20 \times (-2,322)) + \\
&\quad (0,20 \times (-2,322))) \\
&= -(-0,4422 + (-0,4644) + (-0,4644)) \\
&= -(-1,371) \\
&= 1,371
\end{aligned}$$

$$\begin{aligned}
E(> 32,85) &= -(p(\text{menengah atas}|semua) \times \\
&\quad \log_2 p(\text{menengah atas}|semua)) \\
&= -\left(\frac{1}{1} \times \log_2 \frac{1}{1}\right) \\
&= -(1 \times 0) \\
&= 0
\end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{semua}, 16,85) &= 1,4588 - \left(\left(\frac{1}{6} \right) \times 0 \right) + \left(\left(\frac{5}{6} \right) \times 1,371 \right) \\ &= 1,4588 - 0 + 1,1424 \\ &= 0,3164 \end{aligned}$$

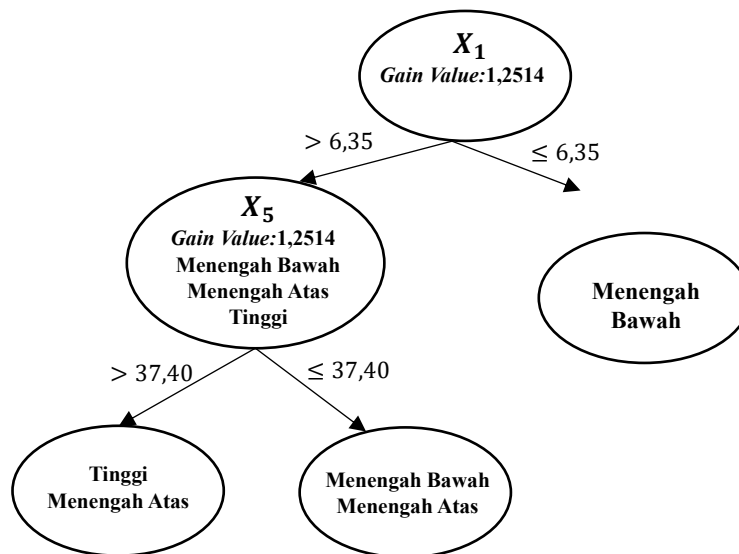
$$\begin{aligned} \text{Gain}(\text{semua}, 30,8) &= 1,4588 - \left(\left(\frac{4}{6} \right) \times 0,811 \right) + \left(\left(\frac{2}{6} \right) \times 1 \right) \\ &= 1,4588 - 0,5407 + 0,3333 \\ &= 1,2514 \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{semua}, 32,85) &= 1,4588 - \left(\left(\frac{5}{6} \right) \times 1,371 \right) + \left(\left(\frac{1}{6} \right) \times 0 \right) \\ &= 1,4588 - 1,1424 + 0 \\ &= 0,3164 \end{aligned}$$

Tabel 3.12 Entropy dan Gain Value X_6

Mean	Jumlah Kasus	Entropy	Gain
16,85	\leq	1	0
	$>$	5	1,371
30,8	\leq	4	0,811
	$>$	2	1
32,85	\leq	5	1,371
	$>$	1	0

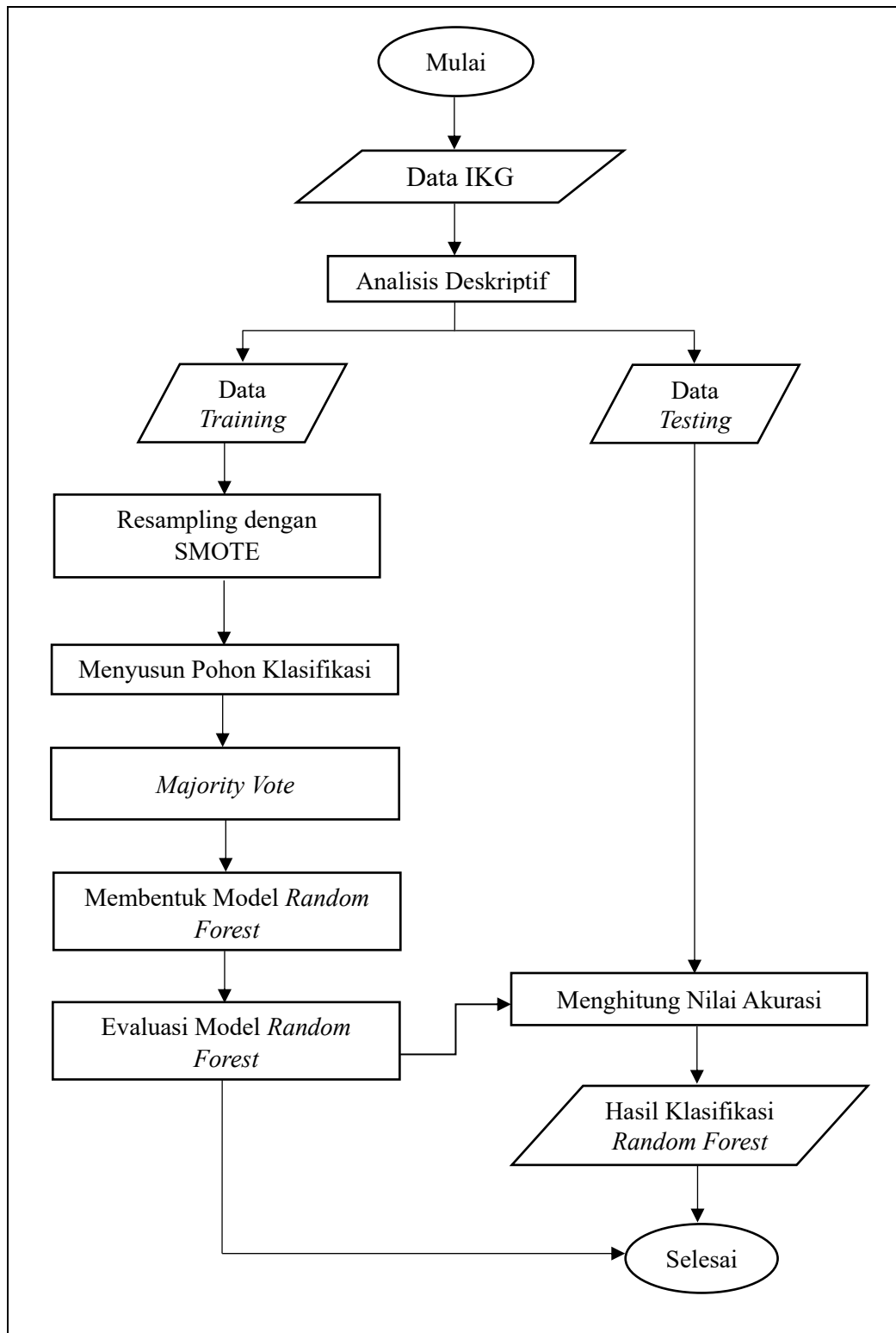
Berdasarkan perhitungan di atas, nilai *information gain* tertinggi terdapat pada variabel X_5 yaitu persalinan tidak di fasilitas kesehatan dengan nilai sebesar 2,6013. Dengan demikian, maka variabel X_5 menjadi *internal node* di bawah *root node* X_1 . Setelah dilakukan perhitungan nilai *entropy* dan nilai *information gain*, kemudian dapat dibentuk pohon keputusan. Berikut merupakan ilustrasi pohon yang dihasilkan dari perhitungan manual.



Gambar 3.1 Pohon Pertama Contoh Data

- c. Melakukan prediksi model dari hasil pohon klasifikasi.
 - d. Mengulangi proses a – c sebanyak n kali hingga diperoleh jumlah pohon klasifikasi yang ditentukan.
 - e. Melakukan *majority voting* untuk estimasi prediksi akhir model klasifikasi.
6. Melakukan uji akurasi terhadap model random forest dengan data testing menggunakan confusion matrix.
 7. Melakukan analisis terhadap hasil klasifikasi yang telah terbentuk dari hasil perhitungan yang telah dilakukan.

3.5 Diagram Alir Penelitian



Gambar 3.2 Diagram Alir Penelitian

BAB IV PEMBAHASAN

4.1 Analisis Statistik Deskriptif

Data pada penelitian ini berjumlah 514 data IKG berdasarkan kabupaten/kota di Indonesia tahun 2021. Semakin rendah skor IKG kabupaten/kota maka kesetaraan gender pada wilayah tersebut akan semakin baik pula. Skor IKG kelas rendah yaitu kurang dari 0,399. Jika IKG berada pada rentang 0,400 – 0,449, maka skor IKG berada pada kelas menengah bawah. Apabila IKG berada pada rentang 0,450 – 0,499 berarti bahwa IKG berada di kelas menengah atas dan ketika skor IKG lebih dari sama dengan 0,500 menunjukkan bahwa wilayah tersebut memiliki kategori IKG di kelas tinggi.



Gambar 4.1 Pie Chart IKG

Gambar 4.1 menunjukkan *pie chart* dari indeks ketimpangan gender pada kabupaten/kota di Indonesia pada tahun 2021. Pada *pie chart* tersebut

memperlihatkan terdapat 245 wilayah atau sebesar 48% yang termasuk ke dalam kategori IKG kelas rendah, 15 % atau 75 wilayah berada pada IKG kelas menengah bawah, 64 wilayah atau sebanyak 12% termasuk dalam IKG kelas menengah atas dan 25% lainnya yaitu sebanyak 130 wilayah termasuk dalam kategori kelas tinggi. Kategori kelas pada nilai indeks ketimpangan gender dipengaruhi oleh beberapa indikator yang kemudian menjadi variabel independen dalam penelitian ini, yaitu faktor risiko kematian ibu (X_1), fertilitas remaja (X_2), parlemen pada laki-laki (X_3), parlemen pada perempuan (X_4), pendidikan pada laki-laki (X_5), pendidikan pada perempuan (X_6), tingkat partisipasi angkatan kerja laki-laki (X_7) dan tingkat partisipasi angkatan kerja perempuan (X_8). Indikator pertama yang dapat memengaruhi skor indeks ketimpangan gender adalah proporsi persalinan yang tidak dilakukan di fasilitas kesehatan. Berikut merupakan analisis deskriptif berdasarkan kategori kelas pada indikator pertama yaitu X_1 .

Tabel 4.1 Analisis Deskriptif Variabel Persalinan Tidak di Fasilitas Kesehatan

Variabel	Mean	Std. Deviasi	Minimum	Maksimum
Rendah	0,163	0,186	0,001	0,25
Menengah Bawah	0,137	0,071	0,044	0,379
Menengah Atas	0,217	0,100	0,043	0,524
Tinggi	0,388	0,208	0,003	1,000

Berdasarkan Tabel 4.1 mengenai indikator fasilitas kesehatan diketahui bahwa nilai rata-rata proporsi X_1 pada kategori kelas rendah sebesar 0,163 dengan nilai standar deviasi sebesar 0,186, serta nilai minimum dan maksimum masing-masing adalah 0,001 dan 0,250. Indikator fasilitas kesehatan kelas menengah bawah ditunjukkan nilai rata-rata sebesar 0,137 dengan standar deviasinya adalah 0,071 serta nilai minimum dan maksimum secara berurutan sebesar 0,044 dan 0,379. Sementara pada indikator fasilitas kesehatan kelas menengah atas memiliki nilai

rata-rata sebesar 0,217 dengan standar deviasi sebesar 0,100 serta nilai minimum dan maksimum masing-masing adalah 0,043 dan 0,524. Sedangkan indikator fasilitas kesehatan pada kelas tinggi diperoleh nilai rata-rata sebesar 0,388 dengan standar deviasi yang bernilai 0,208 serta nilai minimum dan maksimum secara berurutan adalah 0,003 dan 1,000.

Indikator yang ke dua adalah proporsi wanita usia 15 – 49 tahun yang pernah kawin dan saat melahirkan bayi hidup pertama berusia kurang dari 20 tahun. Berikut perhitungan statistika deskriptif pada indikator X_2 berdasarkan kategori kelas.

Tabel 4.2 Analisis Deskriptif Variabel Wanita Berusia 15 – 49 Tahun yang Pernah Kawin dan Saat Melahirkan Hidup Pertama Berusia Kurang dari 20 Tahun

Variabel	Mean	Std. Deviasi	Minimum	Maksimum
Rendah	0,266	0,096	0,037	0,440
Menengah Bawah	0,279	0,080	0,129	0,540
Menengah Atas	0,311	0,073	0,142	0,440
Tinggi	0,323	0,084	0,082	0,487

Tabel 4.2 merupakan hasil analisis statistika deskriptif pada indikator X_2 yang menunjukkan bahwa nilai rata-rata pada kelas rendah adalah 0,266 dengan standar deviasi sebesar 0,096 serta nilai minimum dan maksimum secara berurutan adalah 0,037 dan 0,440. Pada kelas menengah bawah diperoleh nilai *mean* sebesar 0,279 dengan standar deviasi 0,080 serta nilai minimum dan maksimum masing-masing adalah 0,129 dan 0,540. Sedangkan di kelas menengah atas nilai rata-rata yang diperoleh adalah 0,311 dengan standar deviasi sebesar 0,073 serta skor minimum dan maksimum secara berurutan adalah 0,142 dan 0,440. Kemudian pada kelas yang terakhir yaitu kelas tinggi nilai mean yang diperoleh adalah 0,323 dengan standar deviasi sebesar 0,084 serta nilai minimum dan maksimum masing-masing adalah 0,082 dan 0,487.

Tabel 4.3 Analisis Deskriptif Variabel Laki-laki dan Perempuan yang Duduk di Parlemen

Variabel		Mean	Std. Deviasi	Minimum	Maksimum
Rendah	Laki-laki	81,94	8,19	51,43	96,67
	Perempuan	18,06	8,19	3,33	48,57
Menengah Bawah	Laki-laki	84,25	7,93	66,67	97,14
	Perempuan	15,75	7,93	2,86	33,33
Menengah Atas	Laki-laki	84,80	6,94	68,00	97,50
	Perempuan	15,20	6,94	2,50	32,00
Tinggi	Laki-laki	90,65	8,58	52,00	100,00
	Perempuan	9,35	8,58	0,00	48,00

Berdasarkan Tabel 4.3 secara keseluruhan dapat dilihat bahwa nilai standar deviasi pada masing-masing kelas pada gender laki-laki dan perempuan pada indikator X_3 dan X_4 memiliki hasil nilai yang sama. Sebagai contoh pada kelas rendah, nilai standar deviasi keduanya berada di angka 8,19. Kemudian untuk nilai rata-rata, nilai minimum dan nilai maksimum gender laki-laki cenderung memiliki skor yang lebih tinggi jika dibandingkan dengan perempuan. Sebagai contoh pada kelas tinggi, hasil skor rata-rata laki-laki adalah 90,65 sedangkan perempuan memiliki nilai sebesar 9,35.

Tabel 4.4 Analisis Deskriptif Variabel Penduduk Perempuan dengan Pendidikan Minimal SMA

Variabel		Mean	Std. Deviasi	Minimum	Maksimum
Rendah	Laki-laki	45,96	15,56	18,56	80,95
	Perempuan	41,86	15,45	15,78	81,78
Menengah Bawah	Laki-laki	41,04	12,74	21,29	73,60
	Perempuan	36,07	11,79	16,81	73,95
Menengah Atas	Laki-laki	37,75	10,08	19,06	72,89
	Perempuan	32,49	8,96	16,62	64,71
Tinggi	Laki-laki	36,30	9,81	15,51	65,74
	Perempuan	29,12	9,25	8,30	55,85

Menurut Tabel 4.4 di atas pada masing-masing kelas dan gender untuk indikator X_5 dan X_6 secara keseluruhan nilai rata-rata, standar deviasi, nilai minimum dan nilai maksimum memiliki angka yang cenderung relatif sama atau

tidak memiliki perbedaan signifikan. Sebagai contoh pada kategori kelas menengah bawah, nilai maksimum laki-laki berada di angka 73,60 sedangkan perempuan berada di angka 73,95. Angka ini memiliki selisih yang kecil yaitu 0,35.

Tabel 4.5 Analisis Deskriptif Variabel Angkatan Kerja Terhadap Penduduk Usia Kerja Laki-laki dan Perempuan

Variabel		Mean	Std. Deviasi	Minimum	Maksimum
Rendah	Laki-laki	81,61	3,96	71,68	92,83
	Perempuan	55,56	8,85	36,11	83,82
Menengah Bawah	Laki-laki	83,40	4,39	73,91	95,56
	Perempuan	56,12	10,08	39,24	96,17
Menengah Atas	Laki-laki	83,68	3,18	74,12	89,66
	Perempuan	53,59	9,24	36,79	80,16
Tinggi	Laki-laki	84,48	4,51	71,73	97,05
	Perempuan	57,28	13,09	30,71	99,00

Berdasarkan Tabel 4.5 nilai maksimum pada indikator X_7 dan X_8 memiliki nilai yang relatif sama pada masing-masing kategori kelas dan gendernya. Hal ini terlihat pada selisih yang dimiliki pada setiap gender pada nilai maksimum. Sebagai contoh pada kategori kelas tinggi, nilai maksimum laki-laki sebesar 97,05 dan perempuan sebesar 99,00. Kemudian pada nilai rata-rata, standar deviasi dan nilai minimum untuk masing-masing kategori kelas memiliki nilai yang relatif berbeda antara gender laki-laki dan perempuan. Contohnya adalah pada nilai rata-rata kategori tinggi laki-laki memiliki skor sebesar 84,48 sedangkan perempuan memiliki skor sebesar 57,28.

4.2 Analisis Data

4.2.1 Penentuan Data *Training* dan Data *Testing*

Proses analisis data dengan metode *random forest* umumnya dilakukan dengan melakukan *splitting data* menjadi data *training* dan data *testing*. Data *training* digunakan untuk melakukan pembentukan model klasifikasi dan data *testing* digunakan untuk melakukan pengujian kebaikan model yang telah

dibentuk. Persentase pembagian data *training* dan data *testing* dalam penelitian ini dibagi menjadi 5 kombinasi, yaitu (70%:30%), (75%:25%), (80%:20%), (85%:15%), dan (90%:10%). Setelah kombinasi persentase antara keduanya ditentukan, kemudian dilakukan pembentukan model terhadap setiap kombinasi persentase data *training* dan data *testing* serta melakukan perhitungan terhadap tingkat ketepatan model klasifikasi. Hasil perhitungan dengan nilai akurasi terbaik akan dipilih karena memiliki skor yang seimbang antara data *training* dan data *testing*.

Tabel 4.6 Data Training dan Testing

No	Persentase Data		Tingkat Akurasi
	<i>Training</i>	<i>Testing</i>	
1	70%	30%	0,8377
2	75%	25%	0,8359
3	80%	20%	0,8155
4	85%	15%	0,8182
5	90%	10%	0,902

Tabel 4.6 memperlihatkan hasil perhitungan dari kombinasi persentase data *training* dan data *testing* pada setiap persentase. Kombinasi persentase 90% data *training* dan persentase 10% data *testing* memiliki nilai akurasi tertinggi yaitu sebesar 0,902. Oleh karena itu, pada langkah analisis selanjutnya menggunakan kombinasi persentase (90% : 10%) dengan jumlah data *training* sebanyak 463 dan data *testing* yang berjumlah 51.

4.2.2 Balancing Data dengan Metode SMOTE

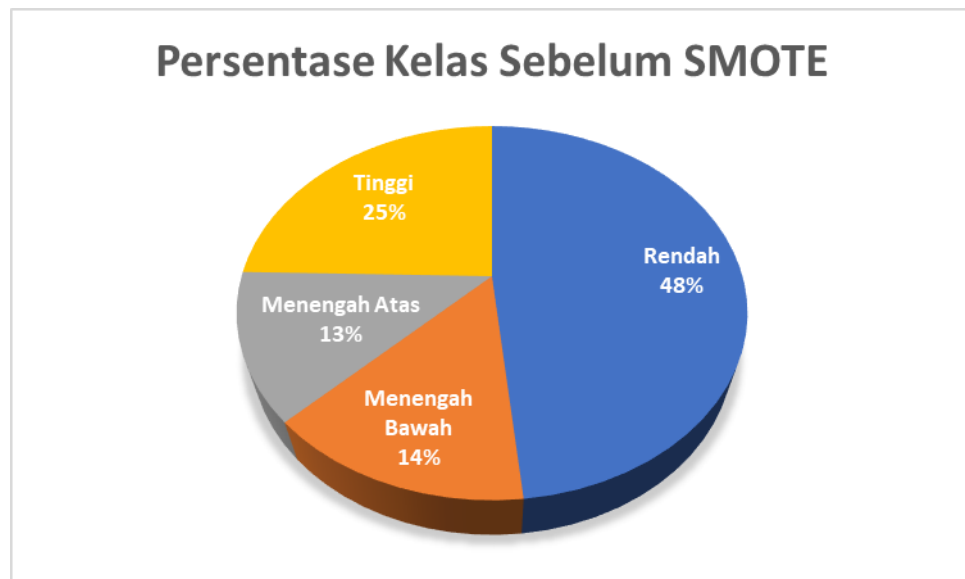
Data *training* pada variabel target penelitian ini memiliki jumlah yang tidak seimbang atau biasa dikatakan sebagai *imbalanced dataset*. Untuk itu, data perlu diseimbangkan agar mencapai hasil yang optimal. Pada penelitian ini sampling data dilakukan dengan menggunakan metode *Synthetic Minority Over-*

sampling Technique (SMOTE). Berikut merupakan jumlah awal sebelum dilakukan SMOTE pada masing-masing kelas.

Tabel 4.7 Frekuensi Data *Training*

Kelas	Frekuensi
Rendah	224
Menengah Bawah	66
Menengah Atas	59
Tinggi	114

Pada Tabel 4.7 ditunjukkan frekuensi dari masing-masing kelas pada data *training* sebelum dilakukan *oversampling* dengan jumlah keseluruhan sebanyak 463 data. Kelas dengan frekuensi terbesar adalah kelas rendah sebanyak 224 dan kelas dengan frekuensi terkecil adalah kelas menengah atas yaitu sebanyak 59. Kelas lainnya yaitu kelas menengah bawah memiliki jumlah sebanyak 66 dan kelas tinggi sebanyak 114. Angka ini menunjukkan celah yang cukup besar.



Gambar 4.2 Jumlah Kelas Sebelum SMOTE

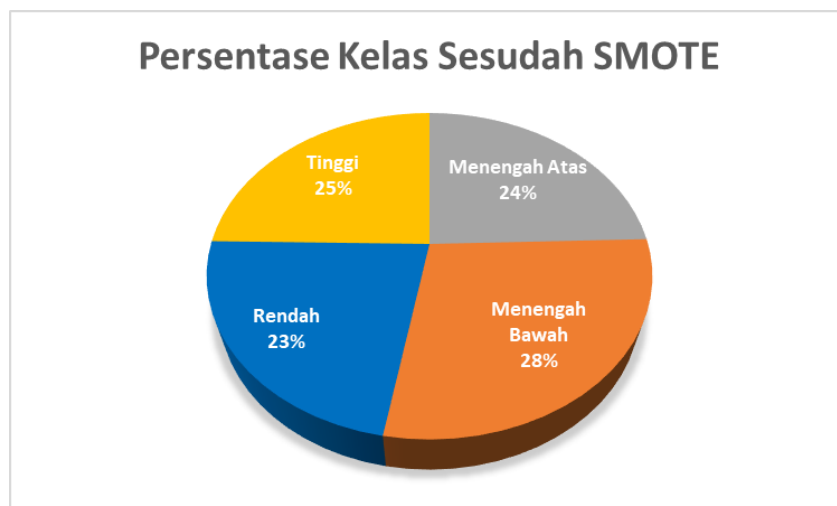
Jika dilihat dengan menggunakan persentase seperti pada Gambar 4.2, maka akan terlihat lebih jelas perbedaan antara kelas mayoritas dengan kelas lainnya. Setelah dilakukan resampling pada data *training* frekuensi dari

keseluruhan kelas menjadi lebih setara. Berikut adalah frekuensi pada masing-masing kelas setelah dilakukan *resampling* menggunakan metode SMOTE.

Tabel 4.8 Frekuensi Setelah *Resampling*

Kelas	Frekuensi
Rendah	224
Menengah Bawah	272
Menengah Atas	236
Tinggi	238

Setelah dilakukan proses *resampling* terhadap data *training*, frekuensi pada setiap kelas berada pada angka yang lebih seimbang. Pada data setelah proses SMOTE, kelas rendah memiliki jumlah sebanyak 224, kelas menengah bawah memiliki jumlah sebanyak 272, kelas menengah atas memiliki jumlah sebanyak 236, dan kelas tinggi sebanyak 238. Kelas minoritas pada data *training* memiliki jumlah yang meningkat dengan menghasilkan sampel sintetis guna mengimbangi frekuensi kelas yang tidak seimbang. Menggunakan cara ini, diharapkan performa model klasifikasi yang dihasilkan menjadi lebih baik mewakili semua kelas.



Gambar 4.3 Jumlah Kelas Setelah SMOTE

Berdasarkan Gambar 4.3, dapat dilihat bahwa persentase dari masing-masing kelas pada data *training* menjadi lebih seimbang. Sebelum dilakukan *resampling* jumlah data sebanyak 463 dan setelah dilakukan *resampling* menjadi 970. Berikut merupakan tabel perbandingan jumlah kelas pada data sebelum dilakukan *resampling* dengan data setelah dilakukan *resampling* menggunakan SMOTE.

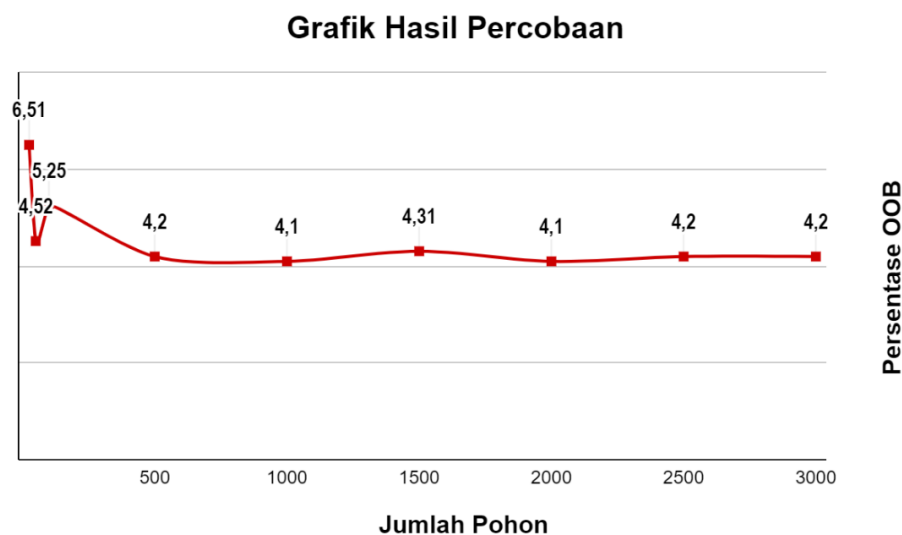
Tabel 4.9 Perbandingan Jumlah Frekuensi

Kelas	Sebelum SMOTE	Sesudah SMOTE
Menengah Bawah	66	272
Menengah Atas	59	236
Tinggi	114	238
Rendah	224	224

Pada tabel 4. terlihat bahwa terdapat tiga kelas minoritas yaitu kelas menengah bawah, kelas menengah atas dan kelas tinggi. Setelah dilakukan SMOTE, ketiga kelas tersebut memiliki jumlah data yang meningkat cukup banyak. Sebelum proses *oversampling*, kelas menengah bawah memiliki jumlah data sebanyak 66, namun setelah proses *resampling* kelas tersebut memiliki jumlah sebanyak 272. Demikian pula yang terjadi pada kelas menengah atas yang mulanya hanya memiliki data sebanyak 59 dan kemudian setelah SMOTE bertambah hingga 236 data. Sedangkan kelas tinggi yang awalnya berjumlah 114 kemudian menjadi 238 data setelah proses SMOTE. Kelas rendah memiliki jumlah yang tetap karena saat ia termasuk ke dalam kelas mayoritas ketika belum dilakukan proses *resampling*. Dengan demikian, proses *oversampling* dengan menggunakan SMOTE telah mencapai tujuannya yaitu *balancing* data pada kelas-kelas minoritas.

4.2.3 Hasil Klasifikasi Menggunakan Metode *Random Forest*

Setelah dilakukan penyeimbangan data pada *data training* dengan menggunakan metode *Synthetic Minority Over-sampling Technique (SMOTE)*, kemudian akan dilakukan analisis klasifikasi dengan menggunakan metode *random forest*. Pada proses klasifikasi indeks ketimpangan gender menggunakan metode *random forest* dilakukan dengan bantuan *software RStudio*. Langkah awal yang dilakukan yaitu menentukan jumlah terbaik dari variabel yang akan digunakan untuk membangun *decision tree* dan jumlah pohon optimal yang akan dibentuk. Penelitian ini akan menggunakan dua variabel terpilih dengan jumlah pohon sebanyak 500 pohon. Jumlah pohon ini dipilih berdasarkan uji coba yang telah dilakukan untuk mengetahui pengaruh jumlah pohon terhadap model klasifikasi yang dihasilkan.



Gambar 4.4 Grafik Hasil Percobaan *n tree*

Menurut penelitian yang dilakukan oleh Oshiro dkk., (2012), *n tree* dengan jumlah yang besar dalam sebuah *forest* hanya mengakibatkan meningkatnya kompleksitas dalam proses komputasi dan juga tidak meningkatkan performa

secara signifikan. Pada observasi lain yang dilakukan oleh mereka pada penelitian, diperlihatkan bahwa tidak terdapat perbedaan secara signifikan pada hasil kinerja antara menggunakan jumlah *n*tree ukuran awal dengan jumlah dua kali lebih banyak dari jumlah aslinya. Hal ini sesuai dengan uji coba yang telah dilakukan pada Gambar 4.3. Berdasarkan Gambar 4.3, percobaan dilakukan terhadap sembilan jumlah pohon (*n*tree) yang berbeda. Percobaan pertama hingga ketiga, hasil pada grafik masih menunjukkan angka yang fluktuatif. Mulai percobaan keempat hingga pada percobaan terakhir, hasil dari *out-of-bag* berada pada angka yang cenderung stabil dengan nilai rata-rata sebesar empat. Jumlah pohon dalam *ensemble learning* seperti metode klasifikasi *random forest* dapat dipengaruhi oleh *strong law of large numbers* dan *tree structurer*. Semakin banyak jumlah pohon yang terbentuk dalam model klasifikasi, maka akan menyebabkan nilai estimasi dari *generalization error* atau nilai *out-of-bag errors* menjadi konvergen (Erkamim dkk., 2023). Maka dari itu, untuk penelitian ini akan menggunakan *n*tree sebanyak 500 pohon agar dapat mengurangi biaya dalam proses komputasi yang dijalankan dengan hasil model klasifikasi yang optimal.

Proses pembangunan pohon dengan menggunakan *bootstrapping sampling* pada data *training* dilakukan hingga didapatkan jumlah pohon yang disesuaikan. Setelah semua pohon dibangun, kemudian diberikan pengondisian sehingga proses klasifikasi pada pohon-pohon tersebut dapat dilakukan dengan *majority voting*. Pada *random forest*, *majority voting* dilakukan dengan melakukan perhitungan dengan jumlah kelas terbanyak dengan menggabungkan semua hasil prediksi dari banyak pohon Keputusan untuk memperoleh hasil

prediksi akhir. Pada data asli indeks ketimpangan gender menggunakan 500 pohon diperoleh hasil klasifikasi dengan jumlah 463 sebagai berikut:

Tabel 4.10 Hasil Klasifikasi

Kategori Kelas	Jumlah Kabupaten/Kota
Rendah	219
Menengah Bawah	69
Menengah Atas	60
Tinggi	115

Terdapat 219 kabupaten/kota dengan kategori kelas rendah, diantaranya adalah Aceh Barat Daya, Bangka Barat, Cirebon, Deli Serdang, Enrekang, Kota Jakarta, Kota Kediri, Kab. Ngawi, Minahasa Selatan, Siau Tagulandang Biaro, Tapanuli Tengah dan Kota Malang. 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 menengah bawah berjumlah sebanyak 69. Beberapa diantaranya adalah Gorontalo Utara, Kota Banjarmasin, Kota Tasikmalaya, Manokwari Selatan, Sumba Barat Daya, dan Timor Tengah Selatan. Kabupaten/kota dengan kategori menengah atas memiliki angka terkecil daripada kategori lainnya yaitu 60. Pada kategori ini hanya terdapat empat kabupaten/kota yang berasal dari Pulau Jawa, yaitu Bandung Barat, Kota Serang, Kabupaten Subang, dan Kabupaten Tasikmalaya. 56 lainnya merupakan kabupaten/kota di luar Pulau Jawa seperti Toli-Toli, Tapanuli Selatan, Wakatobi, Ogan Komering Ulu Timur, Labuhan Batu Selatan, dan Kutai Kartanegara. Sedangkan sisanya termasuk ke dalam kategori tinggi yang berjumlah 115. Terdapat empat kabupaten dari Pulau Jawa yang termasuk dalam kategori tinggi, yaitu Bogor, Garut, Sukabumi dan Sumenep. Daerah dengan kategori IKG tinggi memiliki arti bahwa tingkat kesetaraan gender

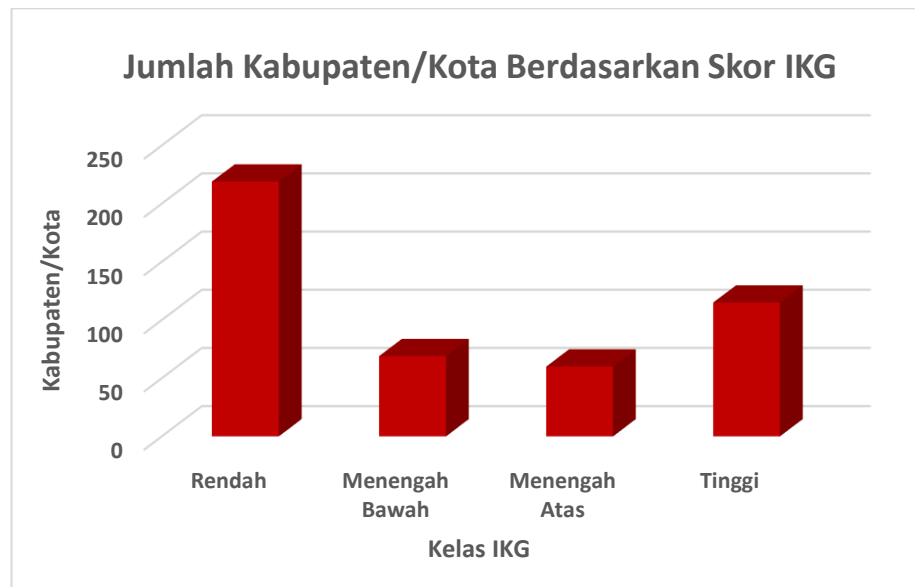
pada wilayah tersebut masih kurang optimal, sehingga pembangunan manusia yang terjadi di daerah tersebut masih terkoreksi dengan ketimpangan gender yang cukup besar.



Gambar 4.5 Sebaran IKG di Indonesia

Berdasarkan sebaran skor indeks ketimpangan gender pada Gambar 4.5, terlihat bahwa sebaran skor IKG terbanyak adalah berwarna biru, yaitu skor rendah. Hal ini dapat menunjukkan bahwa keadaan ketimpangan di Indonesia sudah cukup rendah, meskipun pada beberapa daerah masih terdapat skor IKG dengan kelas tinggi. Salah satunya adalah Kabupaten Sumenep. Pada wilayah Sumenep yang termasuk ke dalam daerah dengan IKG kelas tinggi yaitu sebesar 0,530 poin, berdasarkan data dari pengadilan agama Sumenep terdapat sebanyak 313 dispensasi pernikahan dini yang diajukan pada tahun 2022 (Sanita, 2023). Pernikahan dini pada daerah tersebut disebabkan oleh kepercayaan yang dianut oleh masyarakat. Jika perempuan yang telah berusia 15 hingga 18 tahun namun belum menikah akan menjadi bahan gunjingan masyarakat sekitar dan akan

dilabeli dengan julukan “*sangkal*”, yaitu tidak akan ada laki-laki yang bersedia untuk menikah dengan perempuan tersebut. Karena alasan itulah yang banyak menyebabkan masyarakat Sumenep segera menikahkan anak perempuan mereka (Bawono dkk., 2022).



Gambar 4.6 Grafik Hasil Klasifikasi Indeks Ketimpangan Gender

4.2.4 Evaluasi Model Metode *Random Forest*

Model klasifikasi yang telah diperoleh, kemudian dilakukan evaluasi guna mengetahui kebaikan dari model tersebut. Evaluasi model dilakukan dengan menggunakan *K-Fold Cross Validation* dengan *fold* sebanyak lima. Berikut merupakan hasil perhitungan dengan menggunakan *K-Fold Cross Validation*.

Tabel 4.11 *K-Fold Cross Validation*

<i>Mtry</i>	<i>Accuracy</i>
2	0,9553649
3	0,9468637
4	0,9437277
5	0,9385303
6	0,9353998

Tabel 4.9 menunjukkan bahwa tingkat akurasi terbaik yang diuji dengan menggunakan *5-fold cross validation* yaitu sebesar 0,9553649 dengan menggunakan *mtry* sebanyak dua variabel. Hasil percobaan menunjukkan bahwa jika nilai *mtry* yang digunakan semakin besar, maka tingkat akurasi dari model klasifikasi menjadi lebih homogen. Hal ini dikarenakan ketika nilai *mtry* meningkat, pohon-pohon dalam model klasifikasi akan memiliki lebih banyak *feature* yang sama untuk dijadikan sebagai pertimbangan sehingga menghasilkan keputusan yang mirip antara satu dengan yang lainnya (Mahmuda, 2024).

Tingkat ketepatan hasil klasifikasi IKG pada metode *random forest* dapat diukur menggunakan data *testing* dari *confusion matrix* berikut.

Tabel 4.12 *Confusion Matrix Data Testing*

<i>Confusion Matrix</i>		Kelas Aktual			
		Menengah Atas	Menengah Bawah	Rendah	Tinggi
Kelas Prediksi	Menengah Atas	6	0	0	0
	Menengah Bawah	3	2	1	0
	Rendah	0	1	24	0
	Tinggi	0	0	0	14

Hasil perhitungan *confusion matrix* berdasarkan Tabel 4.10 menggunakan data *testing* IKG diperoleh terdapat enam kabupaten/kota yang memiliki kategori IKG menengah atas diprediksi benar menjadi kabupaten/kota dengan kategori IKG menengah atas, dua kabupaten/kota diprediksi benar memiliki kategori IKG menengah bawah, 24 kabupaten/kota tepat diklasifikasikan sebagai kategori IKG rendah dan 14 kabupaten/kota tepat diklasifikasikan sebagai kategori IKG tinggi. Sedangkan terdapat enam kabupaten/kota yang tidak dapat diklasifikasikan

dengan tepat oleh *random forest*. Kemudian perhitungan nilai *accuracy* pada data *testing* adalah sebagai berikut:

$$\begin{aligned} Accuracy &= \frac{6 + 2 + 24 + 14}{3 + 1 + 1 + 6 + 2 + 24 + 14} \times 100\% \\ &= 90,02\% \end{aligned}$$

Nilai akurasi dari hasil klasifikasi IKG menggunakan metode *random forest* pada data *testing* diperoleh sebesar 90,02%. Angka ini sudah cukup membuktikan bahwa metode *random forest* dapat melakukan analisis data klasifikasi dengan baik. Namun, dari perhitungan *confusion matrix* pada data *testing* menunjukkan bahwa hasil antar kelas memiliki jumlah yang kurang seimbang. Untuk itu perlu dilakukan perhitungan *f1 score* untuk memberikan hasil yang lebih adil mengenai gambaran performa model klasifikasi. Nilai *f1-score* menunjukkan rata-rata harmonis dari nilai *precision* dan *recall*. Berikut merupakan perhitungan secara manual untuk nilai *precision*, *recall* dan *f1-score* pada masing-masing kelas.

1. Kelas Rendah

$$Precision = \frac{24}{(24+1+0+0)} \times 100\%$$

$$= 96\%$$

$$Recall = \frac{24}{(0+1+24+0)} \times 100\%$$

$$= 96\%$$

$$F_1 - Score = 2 \times \frac{(96\% \times 96\%)}{(96\% + 96\%)}$$

$$= 96\%$$

2. Kelas Menengah Bawah

$$Precision = \frac{2}{(2+0+1+0)} \times 100\%$$

$$= 66,67\%$$

$$Recall = \frac{2}{(0+2+1+0)} \times 100\%$$

$$= 66,67\%$$

$$F_1 - Score = 2 \times \frac{(66,67\% \times 66,67\%)}{(66,67\% + 66,67\%)}$$

$$= 66,67\%$$

3. Kelas Menengah Atas

$$Precision = \frac{6}{(6+0)} \times 100\%$$

$$= 100\%$$

$$Recall = \frac{6}{(6+3+0+0)} \times 100\%$$

$$= 66,67\%$$

$$F_1 - Score = 2 \times \frac{(100\% \times 66,67\%)}{(100\% + 66,67\%)}$$

$$= 80\%$$

4. Kelas Tinggi

$$Precision = \frac{14}{(14+0+0+0)} \times 100\%$$

$$= 100\%$$

$$Recall = \frac{14}{(0+0+0+14)} \times 100\%$$

$$= 100\%$$

$$F_1 - Score = 2 \times \frac{(100\% \times 100\%)}{(100\% + 100\%)}$$

$$= 100\%$$

Hasil dari perhitungan nilai *precision*, *recall* dan *f1-score* pada masing-masing kelas menunjukkan pada kelas rendah nilai *precision* sebesar 96% yang berarti bahwa dari prediksi secara keseluruhan yang masuk ke dalam kategori kelas rendah adalah sekitar 96% benar dikategorikan sebagai kelas rendah. Nilai *recall* pada kelas rendah sebesar 96% yang menunjukkan bahwa dari keseluruhan data aktual dengan label kelas rendah, model klasifikasi berhasil melakukan identifikasi sebesar 96% dengan tepat. Nilai *f1-score* pada kelas rendah sebesar 96% mengindikasikan bahwa model klasifikasi dapat mencapai rata-rata harmoni antara nilai *precision* dan nilai *recall* yang berarti bahwa model klasifikasi dengan metode *random forest* mampu dalam melakukan prediksi dengan baik. Pada kelas menengah bawah, nilai *precision* dan nilai *recall* masing-masing memperoleh angka sebesar 66,67%, begitu pula dengan nilai *f1-score*-nya. Hasil nilai *precision*, *recall* dan *f1-score* pada kelas menengah bawah menunjukkan bahwa model dapat melakukan kategorisasi dengan tepat pada kelas tersebut sebesar 66,67%. Kemudian pada kelas menengah atas, nilai *precision* memperoleh nilai sebesar 100%, hal ini memperlihatkan bahwa dari keseluruhan prediksi yang dikategorikan sebagai kelas tersebut adalah benar. Adapun nilai *recall* pada kelas menengah atas memperoleh hasil sebanyak 66,67% yang berarti bahwa dari semua data yang sebenarnya merupakan kelas menengah atas, model mampu melakukan identifikasi sebesar 66,67% secara tepat. Nilai *f1-score* pada kelas ini sebesar 80% yang berarti bahwa rata-rata harmoni antara nilai *precision* dan nilai *recall* memiliki angka yang cukup baik. Kelas terakhir yaitu tinggi dengan nilai *precision*, nilai *recall* dan nilai *f1-score* adalah 100%. Nilai dengan angka tersebut menunjukkan bahwa model klasifikasi sangat mampu dalam melakukan

klasifikasi pada kelas tinggi. Setelah dilakukan perhitungan terhadap masing-masing kelas, kemudian melakukan perhitungan terhadap *f1-score* secara keseluruhan, berikut merupakan perhitungan *f1-score* secara keseluruhan.

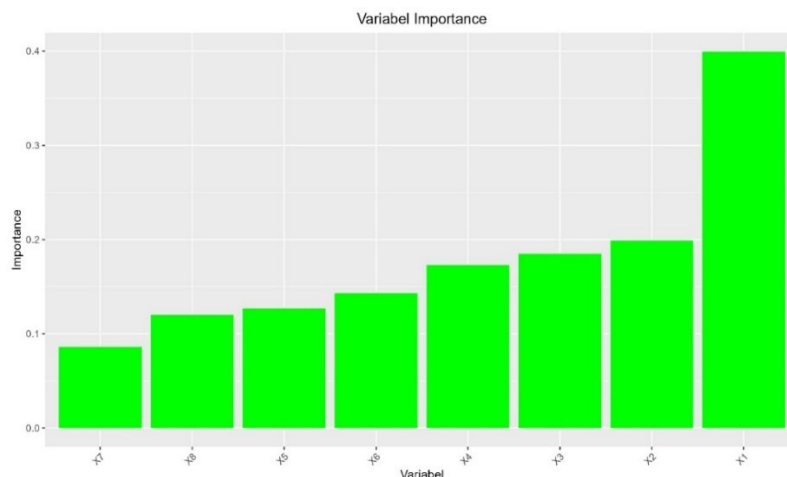
$$\begin{aligned}
 \text{Weighted } F_1 \text{ Score} &= [(0,1176 \times 0,8) + (0,1765 \times 0,4444) + \\
 &\quad (0,4902 \times 0,96) + (0,2745 \times 1)] \times 100\% \\
 &= (0,09408) + (0,0783868) + (0,470592) + \\
 &\quad (0,2745) \times 100\% \\
 &= 91,76\%
 \end{aligned}$$

Berdasarkan hasil pada nilai *weighted f1-score* pada angka 91,76% menunjukkan bahwa model klasifikasi memiliki performa yang baik dalam melakukan klasifikasi untuk semua kelas yang terdapat pada data indeks ketimpangan gender.

Tabel 4.13 Tingkat Ketepatan Model

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Weighted F1-Score</i>
Rendah	96%	96%	96%	91,76%
Menengah Bawah	66,67%	66,67%	66,67%	
Menengah Atas	100%	66,67%	80%	
Tinggi	100%	100%	100%	

Pada analisis klasifikasi indeks ketimpangan gender dengan metode *random forest* dapat diketahui variabel-variabel yang memiliki pengaruh signifikan terhadap proses klasifikasi atau juga dapat disebut sebagai *variable importance*. *Variable importance* ditunjukkan melalui tingkat dari masing-masing variabel seperti yang dapat dilihat pada *plot* berikut.



Gambar 4.7 Variable Importance IKG

Pada Gambar 4.6 dapat dilihat secara eksplisit bahwa variabel X_1 memiliki *bar* tertinggi jika dibandingkan dengan variabel lainnya. Hal ini menunjukkan bahwa X_1 yaitu persalinan yang tidak dilakukan di fasilitas kesehatan memiliki pengaruh dengan tingkat kepentingan paling tinggi selama proses klasifikasi indeks ketimpangan gender menggunakan metode *random forest*. Hal ini sesuai dengan fakta yang ada di lapangan. 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. Sesuai dengan penelitian yang dilakukan oleh (Hrp & Nst, 2021) dari 56 responden, terdapat 10 orang yang memilih melakukan persalinan di fasilitas kesehatan dan 46 orang memilih melakukan persalinan di fasilitas non-kesehatan. atau di rumah karena tidak ada dukungan dari pihak keluarga dalam

memilih tempat persalinan sehingga lebih mengutamakan untuk mengeluarkan biaya lebih sedikit.

Kemudian variabel X_2 adalah wanita berusia 15 – 49 tahun yang pernah kawin dan saat melahirkan hidup pertama berusia kurang dari 20 tahun dengan tingkat kepentingan kedua setelah variabel X_1 . Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) menyebutkan bahwa angka kelahiran pada remaja dengan usia 15 – 19 tahun di Indonesia mengalami kenaikan pada tahun 2022 sebanyak enam poin. Menurut penelitian yang dilakukan oleh Devi dkk., (2022) fenomena ibu melahirkan pada usia remaja banyak ditemui di daerah pedesaan.

Variabel X_3 (laki-laki yang duduk di parlemen) dan X_4 (perempuan yang duduk di parlemen) memiliki tingkat kepentingan yang relatif sama. Berdasarkan Keputusan KPU Nomor 1562 Tahun 2023 menyebutkan bahwa terdapat sebanyak 9.917 pendaftar tetap untuk kursi DPR RI dengan 37,13% atau sebanyak 3.676 pendaftar di dalamnya adalah berjenis kelamin perempuan. Kemudian berdasarkan data Badan Pusat Statistik pada tahun 2022 menunjukkan bahwa sebanyak 21,74% perempuan terlibat dalam urusan parlemen di Indonesia. Hal ini berarti bahwa sebanyak 78,26% urusan parlemen dilakukan oleh laki-laki.

Variabel X_5 (penduduk laki-laki dengan pendidikan minimal SMA) dan X_6 (penduduk perempuan dengan pendidikan minimal SMA) memiliki tingkat kepentingan yang cukup rendah. Jika dilihat berdasarkan data laporan survey sosial nasional (Susenas) yang diterbitkan oleh BPS, pada tahun 2023 sebanyak 37,60% perempuan memiliki pendidikan minimal SMA ke atas dan 42,62%

lainnya adalah laki-laki. Angka ini cukup menunjukkan bahwa rasio antara laki-laki dan perempuan hampir seimbang.

Begitu juga dengan variabel X_7 (angkatan kerja terhadap penduduk usia kerja (laki-laki) dan X_8 (angkatan kerja terhadap penduduk usia kerja perempuan) yang memiliki tingkat kepentingan yang hampir sama satu dengan yang lainnya. Namun, variabel X_7 memiliki tingkat kepentingan paling rendah yang berarti bahwa variabel tersebut tidak berpengaruh secara signifikan terhadap pembentukan model klasifikasi. Tingkat partisipasi angkatan kerja pada tahun 2021 berdasarkan jenis kelamin, terdapat 86,70% laki-laki dan 58,84% perempuan tenaga kerja. Jika dilihat trend dari tahun ke tahun, persentase tingkat partisipasi angkatan kerja perempuan terdapat peningkatan, seperti pada tahun 2023 yang berada di angka 60,18% dan laki-laki sebesar 86,97%.

4.3 Kajian Penelitian dalam Perspektif Islam

Dimensi pasar tenaga kerja merupakan salah satu dimensi pembentuk indeks ketimpangan gender yang memuat satu indikator yaitu tingkat partisipasi angkatan kerja laki-laki dan perempuan. Tingkat partisipasi angkatan kerja adalah angka persentase penduduk di suatu wilayah dengan usia 15 tahun ke atas yang merupakan termasuk dalam angkatan kerja, berkaitan dengan ini terdiri atas laki-laki dan perempuan. Skor tingkat partisipasi angkatan kerja menunjukkan jumlah penduduk usia kerja yang aktif bekerja atau aktif secara ekonomi dalam suatu wilayah. Nilai tingkat partisipasi angkatan kerja yang tinggi menunjukkan bahwa pasokan tenaga kerja yang ada untuk melakukan produksi barang dan jasa di perekonomian yang berbanding lurus. Bekerja memiliki tujuan yang berbeda bagi setiap orang, diantaranya adalah bekerja untuk memperoleh *ridla* Allah SWT, bekerja untuk

memenuhi kebutuhan sehari-hari, bekerja untuk sedekah, dan bekerja untuk memperoleh nafkah (Kurniawan, 2019).

Berdasarkan QS. At-Taubah:105, umat manusia di seluruh dunia ini diperintahkan oleh Allah SWT untuk melakukan pekerjaan yang memiliki manfaat bagi diri sendiri dan juga bagi orang lain. Semua pekerjaan yang dilakukan oleh umat manusia akan selalu dilihat oleh Allah, Rasul serta orang-orang mukmin yang akan diperlihatkan oleh Allah SWT saat hari kiamat di kemudian hari. Saat hari kiamat tiba, umat manusia yang senantiasa melakukan pekerjaan dengan niat baik akan diberikan balasan sesuai dengan perbuatannya masing-masing.

Menurut M. Quraish Shihab, ayat ini memiliki tujuan agar umat manusia dapat saling mendorong untuk melakukan introspeksi kepada diri sendiri dalam hal pekerjaan yang dilakukan bahwa setiap perbuatan yang dilakukan akan senantiasa dilihat oleh Allah dan diberi balasan sesuai dengan porsinya masing-masing (Kurniawan, 2019).

Berdasarkan paparan makna di atas, bekerja dapat dilakukan oleh laki-laki maupun perempuan. Bekerja tidak memiliki *gender boundaries* dan tidak diistimewakan hanya untuk salah satu gender karena sifatnya adil. Sifat adil diungkapkan dalam QS. Al-Maidah:08. Berdasarkan ayat tersebut, terdapat kata keadilan, yang mana ini merupakan salah satu sifat yang sangat penting untuk diterapkan dalam kehidupan sehari-hari, salah satunya dalam kehidupan kerja. Berperilaku secara adil dapat berpengaruh dalam lingkungan sekitarnya, yang tentu saja hal ini juga dapat berlaku dalam lingkungan pekerjaan. Kata “*syahadah*” dalam ayat tersebut memiliki makna perumpamaan agar umat manusia memiliki sikap adil dan sikap berani dalam menunjukkan kebenaran layaknya tugas seorang hakim

ketika sidang berlangsung. Segala sesuatu yang akan dan sedang dikerjakan wajib dilakukan karena Allah karena Allah melarang untuk memiliki kebencian atau permusuhan terhadap suatu kaum. Dapat disimpulkan bahwasanya antara laki-laki dan perempuan tidak diperbolehkan untuk bersikap secara tidak adil, khususnya ketika dalam beban pekerjaan (Ma'arif & Insiyah, 2017). Meskipun begitu, dalam islam mengakui adanya perbedaan antara laki-laki dan perempuan. Namun hal itu bukanlah perbedaan dalam konteks diskriminasi gender. Perbedaan ini justru dibuat dengan tujuan agar terbentuknya hubungan harmonis dengan dasar rasa kasih sayang atau *mawaddah wa rahmah*. Kualitas seseorang di hadapan Allah tidak dibedakan berdasarkan jenis kelamin manusia karena di hadapan Allah semua makhluk-Nya memiliki potensi yang sama besar .

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 dengan metode *Random Forest* menunjukkan model terbaik menggunakan parameter $n_{tree} = 500$ dan $m_{try} = 2$ dengan perbandingan data *training* 90% dan data *testing* 10%. Pada hasil klasifikasi didapatkan 219 Kabupaten/Kota dengan kategori indeks ketimpangan gender rendah, 69 Kabupaten/Kota diklasifikasikan dengan skor indeks ketimpangan gender menengah bawah, 60 Kabupaten/Kota diklasifikasikan sebagai wilayah dengan skor indeks ketimpangan gender menengah atas, dan 115 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 *random forest*.
2. Tingkat akurasi pada hasil klasifikasi indeks ketimpangan gender pada 514 Kabupaten/Kota Indonesia menggunakan metode klasifikasi *random forest*

pada data *testing* diperoleh nilai *accuracy* sebesar 90,02%, hal ini berarti model yang telah terbentuk mampu melakukan klasifikasi dengan tingkat keakuratan sebesar 90,02%. Selain nilai *accuracy*, diperoleh juga nilai dari *f1-score* sebesar 91,76%.

5.2. Saran

Sebagai bentuk upaya dalam pengembangan serta penyempurnaan penelitian berkesinambungan, oleh karena itu penulis memberikan saran sebagai berikut:

1. Pada penelitian selanjutnya dapat dilakukan dengan metode klasifikasi yang berbeda sehingga dapat dilakukan perbandingan tingkat akurasi antar metode klasifikasi.
2. Pada penelitian selanjutnya dapat dilakukan dengan menggunakan kombinasi data *training* dan *testing* yang berbeda dalam pembagian *dataset*.
3. Pada penelitian selanjutnya disarankan untuk melakukan evaluasi lebih mendalam terhadap hasil model klasifikasi dari sumber eksternal untuk memberikan hasil yang lebih akurat guna pengembangan bidang yang diteliti.

DAFTAR PUSTAKA

- Adam, A., Haddade, H., & Damis, D. R. (2022). Wawasan Al-Quran Tentang Kesetaraan Gender. Dalam *Wawasan Al-Quran Tentang Kesetaraan Gender* (Vol. 8, Nomor 2).
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme Pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2), 50–58. <https://doi.org/10.23960/Komputasi.V8i2.2615>
- Alonso, S., Marques, G., Agarwal, D., Díez, I. D. La T., & Franco-Martín, M. (2022). Comparison Of Machine Learning Algorithms in The Prediction of Hospitalized Patients with Schizophrenia. *Sensors*, 22(7). <https://doi.org/10.3390/S22072517>
- Apriliandra, S., & Krisnani, H. (2018). Perilaku Diskriminatif Pada Perempuan Akibat Kuatnya Budaya Patriarki di Indonesia Ditinjau Dari Perspektif Konflik. *Jurnal Kolaborasi Resolusi Konflik*, 3(1), 1–13.
- Apriliandra, S., & Krisnani, H. (2020). Perilaku Diskriminatif Pada Perempuan Akibat Kuatnya Budaya Patriarki di Indonesia Ditinjau Dari Perspektif Konflik. *Jurnal Kolaborasi Resolusi Konflik*, 3(1), 1–13.
- Bawono, Y., Setyaningsih, Hanim, L. M., Masrifah, & Astuti, J. S. (2022). Budaya Dan Pernikahan Dini Di Indonesia. *Jurnal Dinamika Sosial Budaya*, 24(1). <https://doi.org/10.26623/Jdsb.V21i2.1698>
- Bps. (2022). *Kajian Perhitungan Indeks Ketimpangan Gender 2022*.
- Breiman, L. (2001). *Random Forests* (Vol. 45).
- Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic Minority Over-Sampling Technique. Dalam *Journal of Artificial Intelligence Research* (Vol. 16).
- Damayanti, D. R., & Purwinarko, A. (2024). Application Of C4.5 Algorithm Using Synthetic Minority Oversampling Technique (Smote) And Particle Swarm Optimization (Pso) For Diabetes Prediction. *Recursive Journal of Informatics*, 2(1). <https://doi.org/10.15294/Rji.V2i1.64928>
- Deris, L. R. V., Bhinadi, A., & Nuryadin, D. (2022). Pengaruh Ketimpangan Gender Terhadap Pertumbuhan Ekonomi di Indonesia (34 Provinsi) Tahun 2015-2020. *Sibatik Journal: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, 1(12), 2947–2958. <https://doi.org/10.54443/Sibatik.V1i12.481>
- Devi, Y. P., Ekoriano, M., Sari, D. P., & Muthmainnah, M. (2022). Factors Associated with Adolescent Birth In Indonesia; A National Survey. *Rural And Remote Health*, 22(2). <https://doi.org/10.22605/Rrh7219>

- Erkamim, M., Suswadi, Subarkah, M. Z., & Widarti, E. (2023). Komparasi Algoritme Random Forest Dan Xgboosting Dalam Klasifikasi Performa Umkm. *Jurnal Sistem Informasi Bisnis*, 13(2), 127–134. <https://doi.org/10.21456/Vol13iss2pp127-134>
- Fuleky, P. (2020). *Macroeconomic Forecasting in The Era Of Big Data*. <http://www.springer.com/series/5667>
- Hasibuan, L. (2006). *Antara Emansipasi Dan Peran Ganda Perempuan (Analisa Fakta Sosial Terhadap Kasus Ketimpangan Gender)*.
- He, H., Yan, H., & Tan, V. Y. F. (2022). Information-Theoretic Characterization of The Generalization Error for Iterative Semi-Supervised Learning. Dalam *Journal of Machine Learning Research* (Vol. 23). <http://jmlr.org/papers/v23/22-0541.html>.
- Hilmiyah, F. (2017). *Prediksi Kinerja Mahasiswa Menggunakan Support Vector Machine Untuk Pengelola Program Studi Di Perguruan Tinggi (Studi Kasus: Program Studi Magister Statistika Its)*.
- Hrp, I. P., & Nst, I. N. (2021). Faktor-Faktor Yang Mempengaruhi Keputusan Ibu Dalam Pemilihan Tempat Dan Penolong Persalinan Di Desa Rondaman. Dalam *Jurnal Ilmiah Kesehatan Diagnosis* (Vol. 16).
- Karisma, R. D. L., Kormitson, A., & Kuswanto, H. (2015). *Random Forest of Modified Risk Factor on Ischemic and Hemorrhagic*. 2010, 26–41.
- Kemenpppa. (2024). *Laporan Kinerja Instansi Pemerintah*.
- Khotimah, A. C., & Utami, E. (2022). Comparison Naïve Bayes Classifier, K-Nearest Neighbor and Support Vector Machine in The Classification of Individual On Twitter Account. *Jurnal Teknik Informatika (Jutif)*, 3(3). <https://doi.org/10.20884/1.Jutif.2022.3.3.254>
- Kpppa. (2022). *Pembangunan Manusia Berbasis Gender*.
- Kulkarni, V. Y. (2013). Random Forest Classifiers: A Survey and Future Research Directions. Dalam *International Journal of Advanced Computing* (Vol. 36, Nomor 1).
- Kurniawan, R. (2019). Urgensi Bekerja Dalam Al-Qur'an. *Transformatif*, 3(1), 42–67. <https://doi.org/10.23971/Tf.V3i1.1240>
- Loupe, G. (2014). *Random Forest from Theory to Practice*.
- Ma'arif, M. S., & Insiyah, I. (2017). *Ayat-Ayat Penegakan Hukum Perspektif Para Musafir (Studi Tematik Prinsip-Prinsip Implementasi Adil Dalam Al-Qur'an)*.
- Mahmuda, S. (2024). Implementasi Metode Random Forest Pada Kategori Konten Kanal Youtube. *Jurnal Jendela Matematika*, 2.

- Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. *Technologies*, 9(4). <https://doi.org/10.3390/Technologies9040081>
- Nicodemus, K. K. (2011). Letter To the Editor: On the Stability and Ranking of Predictors from Random Forest Variable Importance Measures. Dalam *Briefings in Bioinformatics* (Vol. 12, Nomor 4, Hlm. 369–373). <https://doi.org/10.1093/Bib/Bbr016>
- Nugroho, A., & Religia, Y. (2021). Analisis Optimasi Algoritma Klasifikasi Naive Bayes Menggunakan Genetic Algorithm Dan Bagging. *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, 5(3), 504–510. <https://doi.org/10.29207/Resti.V5i3.3067>
- Nugroho, K., Noersasongko, E., Purwanto, Muljono, Fanani, A. Z., Affandy, & Basuki, R. S. (2019). Improving Random Forest Method to Detect Hatespeech and Offensive Word. *2019 International Conference on Information and Communications Technology, Icoiact 2019, July*, 514–518. <https://doi.org/10.1109/Icoiact46704.2019.8938451>
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How Many Trees in A Random Forest? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7376 *Lnai*, 154–168. https://doi.org/10.1007/978-3-642-31537-4_13
- Parmar, A., Katariya, R., & Patel, V. (2019). A Review on Random Forest: An Ensemble Classifier. *Lecture Notes on Data Engineering and Communications Technologies*, 26, 758–763. https://doi.org/10.1007/978-3-030-03146-6_86
- Rodríguez, J. D., Pérez, A., & Lozano, J. A. (2010). Sensitivity Analysis of K-Fold Cross Validation in Prediction Error Estimation. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 569–575. <https://doi.org/10.1109/TPAMI.2009.187>
- Sakina, A. I., & Siti, D. H. (2017). Menyoroti Budaya Patriarki Di Indonesia. *Social Work Journal*, 7(1), 1–129. <http://www.jurnalperempuan.org/blog2/-Akar-Sanita>
- Sanita, M. (2023, Oktober 11). *Menilik Tingginya Angka Pernikahan Dini Di Sumenep Madura, Alasan Religi Hingga Kurangnya Edukasi?* Universitas Islam Indonesia: Ilmu Komunikasi.
- Sari, R. P., Sarfiah, S. N., & Indrawati, L. R. (2019). Analisis Pengaruh Ketimpangan Gender Terhadap Produk Domestik Regional Bruto (Pdrb) Tahun 2011-2017 (Studi Kasus 6 Kota Di Provinsi Jawa Tengah). *Dinamic: Directory Journal Of ...*, 1, 467–478.
- Sasongko, T. B. (2016). Komparasi Dan Analisis Kinerja Model Algoritma Svm Dan Pso-Svm (Studi Kasus Klasifikasi Jalur Minat Sma). Dalam *Jurnal Teknik Informatika Dan Sistem Informasi* (Vol. 2).

- Sholichah, A. S. (2021). Partisipasi Perempuan Di Masa Nabi Muhammad Dan Implikasinya. *Al Amin: Jurnal Kajian Ilmu Dan Budaya Islam*, 4(01).
- Subki, M., Sugiarto, F., & Sumarlin. (2021). Penafsiran Qs. Al-Hujurat [49] Ayat 13 Tentang Kesetaraan Gender Dalam Al-Qur'an Menurut Quraish Shihab Dan Sayyid Quthb. *Al Furqan: Jurnal Ilmu Al Quran Dan Tafsir*, 4(1), 12–28. <https://doi.org/10.58518/Alfurqon.V4i1.634>
- Syukron, A., Saputro, E., Widodo, P., & Sardiarinto. (2023). Penerapan Metode Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung. Dalam *Jurnal Teknologi Informasi Dan Terapan (J-Tit)* (Vol. 10, Nomor 1). <https://doi.org/10.25047/Jtit.V10i1.312>
- Tibshirani, R. J., & Efron, B. (1993). An Introduction to The Bootstrap. Dalam *Regression Analysis with Applications B.G. Wetherill* (Vol. 17, Nomor 2).
- Uddin, M. T., & Uddin, M. A. (2015). A Guided Random Forest Based Feature Selection Approach for Activity Recognition. *2nd International Conference on Electrical Engineering and Information and Communication Technology, Iceeict 2015, May*, 21–23. <https://doi.org/10.1109/Iceeict.2015.7307376>
- Vera, A. S. Y. (2016). The Impact of Gender Inequality on Economic Growth In Indonesia. *Sosio Informa*, 2(1), 89–101.
- Wibawa, A. P., Guntur, M., Purnama, A., Akbar, M. F., & Dwiyanto, F. A. (2018). *Metode-Metode Klasifikasi*. 3(1), 134–138.

LAMPIRAN

Lampiran 1 Data Indeks Ketimpangan Gender

No.	Wilayah	Kelas	X1	X2	X3	X4	X5	X6	X7	X8
1	Simeulue	Rendah	3,60	22,10	80,00	20,00	52,19	40,58	82,29	59,46
2	Aceh Singkil	Tinggi	27,20	31,10	88,00	12,00	43,75	41,26	82,46	42,85
3	Aceh Selatan	Rendah	6,10	19,50	90,00	10,00	38,57	33,38	77,52	45,14
4	Aceh Tenggara	Menengah Bawah	16,50	23,10	90,00	10,00	54,67	43,91	78,70	60,83
5	Aceh Timur	Menengah Bawah	7,30	24,10	92,50	7,50	33,90	29,36	79,35	39,86
6	Aceh Tengah	Rendah	7,90	25,90	83,33	16,67	48,59	45,16	85,77	66,55
7	Aceh Barat	Menengah Bawah	7,40	19,50	92,00	8,00	49,98	47,18	78,64	40,85
8	Aceh Besar	Menengah Bawah	4,40	18,50	97,14	2,86	53,69	52,86	75,74	46,83
9	Pidie	Rendah	4,70	13,20	82,50	17,50	40,83	34,01	76,03	46,08
.
.
.
508	Lanny Jaya	Tinggi	58,70	16,40	100,00	0,00	44,59	25,60	93,92	97,71
509	Mamberamo Tengah	Menengah Bawah	10,50	15,20	95,00	5,00	52,69	46,32	95,56	96,17
510	Yalimo	Tinggi	45,10	18,90	100,00	0,00	22,71	20,37	93,58	93,80
511	Puncak	Tinggi	29,90	34,50	96,00	4,00	40,17	28,59	91,25	87,40
512	Dogiyai	Tinggi	100,00	33,60	100,00	0,00	30,77	19,24	84,39	72,18
513	Intan Jaya	Tinggi	55,80	48,70	100,00	0,00	40,60	18,93	84,46	67,01
514	Deiyai	Tinggi	63,80	8,20	100,00	0,00	33,45	9,70	90,80	78,86
515	Kota Jayapura	Rendah	10,80	23,20	74,36	25,64	74,51	71,19	75,02	50,70

Lampiran 2 Syntax Model *Random Forest* dengan RStudio

```
# INSTALL AND LOAD LIBRARY
library(smotefamily)
#IMPORT LIBRARY
library(ggplot2)
library(lattice)
library(caret)
library(randomForest)
library(readxl)
library(openxlsx)

# INPUT DATA
DataIKG <- read_excel("DataIKG.xlsx", sheet = "ALL DATA",
col_types = c("text", "text", "numeric", "numeric", "numeric", "numeric",
"numeric", "numeric",
"numeric", "numeric"))
DataIKG$Kelas <- as.factor(DataIKG$Kelas)
head(DataIKG)

# SPLITTING DATA
set.seed(1000)
n <- nrow(DataIKG) # Untuk mendefinisikan jumlah data
ntrain <- round(n * 0.90) # Persentase data training
a <- sample(n, ntrain) # Agar data random
train_ikg <- DataIKG[a,-1] # Membentuk data training
test_ikg <- DataIKG[-a,-1] # Membentuk data testing
```

Lanjutan Lampiran 2 Syntax Model Random Forest dengan RStudio

```
#SMOTE1
ikgsmote1 <- SMOTE(X = train_ikg[,-1],
                  target = train_ikg[,1],
                  dup_size = 1)
ikgsmote1 <- ikgsmote1$data # extract only the balanced dataset
ikgsmote1$class <- as.factor(ikgsmote1$class)
ikgsmote1

# SMOTE2
ikgsmote2 <- SMOTE(X = ikgsmote1[,-9],
                  target = ikgsmote1[,9],
                  dup_size = 1)
ikgsmote2 <- ikgsmote2$data # extract only the balanced dataset
ikgsmote2$class <- as.factor(ikgsmote2$class)
ikgsmote2

# SMOTE3
ikgsmote3 <- SMOTE(X = ikgsmote2[,-9],
                  target = ikgsmote2[,9],
                  dup_size = 1)
ikgsmote3 <- ikgsmote3$data # extract only the balanced dataset
ikgsmote3$class <- as.factor(ikgsmote3$class)
ikgsmote3
```


Lanjutan Lampiran 2 Syntax Model Random Forest dengan RStudio

```
#SMOTE4
ikgsmote4 <- SMOTE(X = ikgsmote3[,-9],
  target = ikgsmote3[,9],
  dup_size = 1)
ikgsmote4<- ikgsmote4$data # extract only the balanced dataset
ikgsmote4$class <- as.factor(ikgsmote4$class)
ikgsmote4

# SMOTE5
ikgsmote5 <- SMOTE(X = ikgsmote4[,-9],
  target = ikgsmote4[,9],
  dup_size = 1)
ikgsmote5<- ikgsmote5$data # extract only the balanced dataset
ikgsmote5$class <- as.factor(ikgsmote5$class)
ikgsmote5

#MENENTUKAN MTRY TERBAIK
set.seed(12345)
oob.values <- vector(length = 10)
for (i in 1:10) {
  temp.model <- randomForest(class ~ ., data=ikgsmote5, mtry = i, ntree = 500 )
  oob.values[i] <- temp.model$serr.rate[nrow(temp.model$serr.rate), 1]
}
```

Lanjutan Lampiran 2 Syntax Model Random Forest dengan RStudio

```
oob.values
min(oob.values)
which(oob.values == min(oob.values))

#MEMBENTUK MODEL RANDOM FOREST
rf <- randomForest(class ~ ., data=ikgsmote5,
                   mtry = 2, importance = TRUE,
                   nodesize = 1, splitrule = "entropy", ntree = 500,
                   proximity = TRUE)
rf

#MELAKUKAN PREDIKSI DAN MENGHITUNG AKURASI DENGAN
CONFUSION MATRIX
#DATA TESTING
b = dplyr::select(test_ikg,-Kelas) #untuk menghilangkan kolom Y
prediction1 = predict(rf, b)
prediction1 #untuk melihat hasil prediksi
c = table(test_ikg$Kelas, prediction1)
d = confusionMatrix(prediction1, test_ikg$Kelas, mode = "everything",
positive="1")
print(d)
```

Lampiran 3 Syntax Evaluasi Model Random Forest dengan RStudio

```
# EVALUASI MODEL DENGAN K-FOLD CROSS VALIDATION

set.seed(12345) # Untuk hasil yang dapat direproduksi

kfold <- trainControl(method = "cv", number = 5)

model_cv <- train(class ~ ., data = ikgsmote5, method = "rf",
                  trControl = kfold)

# MENAMPILKAN HASIL EVALUASI

print(model_cv)
```

Lampiran 4 Syntax Plot Variabel Importance

```
# HITUNG VARIABEL IMPORTANCE

importance_values <- rf$importance

# BUAT BARPLOT VARIABEL IMPORTANCE

importance_df <- data.frame(Variables = rownames(importance_values)[-1],
                           Importance = importance_values[-1,1])

importance_df <- importance_df[order(importance_df$Importance, decreasing
= TRUE), ]

View(importance_df)

barplot <- ggplot(importance_df, aes(x = reorder(Variables, Importance), y =
Importance)) +
  geom_bar(stat = "identity", fill = "pink") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(x = "Variabel", y = "Importance", title = "Variabel Importance")+
  theme(plot.title = element_text(hjust = 0,5))

barplot

# SIMPAN PLOT SEBAGAI FILE JPEG

ggsave("barplot_importance.jpg", plot = barplot, width = 10, height = 6, units
= "in", dpi = 300)
```

Lampiran 5 Output Tingkat Akurasi Model Random Forest

#DATA TESTING

Confusion Matrix and Statistics

	prediction2			
	Menengah Atas	Menengah Bawah	Rendah	Tinggi
Menengah Atas	6	0	0	2
Menengah Bawah	3	2	1	0
Rendah	0	1	24	0
Tinggi	0	0	0	14

Overall Statistics

Accuracy : 0.902
 95% CI : (0,7859, 0,9674)
 No Information Rate : 0,4902
 P-Value [Acc > NIR]: 5,136e-10

Kappa : 0,8507

Mcnemar's Test P-Value : NA

Statistics by Class

	Class: Menengah Atas	Class: Menengah Bawah	Class: Rendah	Class: Tinggi
Sensitivity	0,6667	0,66667	0,9600	1,0000
Specificity	1,0000	0,91667	0,9615	1,0000
Pos Pred Value	1,0000	0,33333	0,9615	1,0000
Neg Pred Value	0,9333	0,97778	0,9615	1,0000
Precision	1,0000	0,33333	0,9600	1,0000
Recall	0,6667	0,66667	0,9600	1,0000
F1	0,8000	0,44444	0,9600	1,0000
Prevelance	0,1765	0,05882	0,4902	0,2745
Detection Rate	0,1176	0,03922	0,4706	0,2745
Detection Prevelance	0,1176	0,11765	0,4902	0,2745
Balanced Accuracy	0,8333	0,79167	0,9608	1,0000

Lampiran 6 Output Evaluasi Model *Random Forest*

#EVALUASI MODEL

Resampling: Cross-Validated (5-fold)

Resampling result across tuning parameters:

mtry	Accuracy	Kappa
2	0,9479053	0,9300383
5	0,9416499	0,9216639
8	0,9322802	0,9090774

Accuracy was used to select the optimal model using the

Largest value.

The final value for the model was mtry = 2.

RIWAYAT HIDUP



Anindya Luthfiani Susetyo, lahir di Cirebon pada 24 November 2001. Penulis merupakan anak pertama dari dua bersaudara dari pasangan Bapak Edi Susetyo dan Ibu Tanri Janah. Penulis telah menempuh pendidikan mulai dari TK Salsabila yang lulus pada tahun 2007, dilanjutkan menempuh pendidikan sekolah dasar di SDN 2 Kartini Cirebon dan lulus pada tahun 2013. Kemudian penulis melanjutkan pendidikan sekolah menengah pertama di MTSN 1 Kota Cirebon dan lulus pada tahun 2016. Selanjutnya menempuh pendidikan sekolah menengah atas di SMAN 1 Jogorogo dan lulus pada tahun 2019. Pada tahun 2020, penulis melanjutkan pendidikan di Universitas Islam Negeri Maulana Malik Ibrahim Malang pada Program Studi Matematika, Fakultas Sains dan Teknologi. Selama menempuh pendidikan tinggi, penulis aktif mengikuti beberapa kegiatan. Penulis bergabung dalam organisasi HMJ “Integral” Matematika selama satu periode sebagai anggota dan koordinator Divisi Pengembangan Minat dan Bakat. Penulis juga pernah menjadi asisten praktikum Statistika Elementer selama satu semester. Selain itu penulis juga mengikuti kegiatan di luar kampus seperti pelatihan dan seminar.



**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 : Anindya Luthfiani Susetyo
NIM : 200601110085
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Penentuan Klasifikasi Indeks Ketimpangan Gender
di Indonesia Menggunakan Metode Random Forest Classifier
Pembimbing I : Ria Dhea Layla Nur Karisma, M.Si.
Pembimbing II : Juhari, M.Si.

No	Tanggal	Hal	Tanda Tangan
1.	20/07/2023	Konsultasi Topik dan Data	1.
2.	18/10/2023	Konsultasi Bab I	2.
3.	01/11/2023	Konsultasi Bab I, II, dan III	3.
4.	01/01/2023	ACC Bab I, II, dan III	4.
5.	29/11/2023	Konsultasi Kajian Agama Bab I dan II	5.
6.	30/11/2023	ACC Kajian Agama Bab I dan II	6.
7.	12/01/2024	ACC Seminar Proposal	7.
8.	15/03/2024	Konsultasi Revisi Seminar Proposal	8.
9.	18/03/2024	Konsultasi Bab IV dan V	9.
10.	22/03/2024	ACC Bab IV dan V	10.
11.	21/03/2024	Konsultasi Kajian Agama Bab IV	11.
12.	25/03/2024	ACC Kajian Agama Bab IV	12.
13.	26/03/2024	ACC Seminar Hasil	13.
14.	19/04/2024	Konsultasi Revisi Seminar Hasil	14.
15.	13/06/2024	ACC Sidang Skripsi	15.



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

16.	21/06/2024	ACC Revisi Akhir	16.
-----	------------	------------------	-----

Malang, 21 Juni 2024

Mengetahui,

Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.

NIP. 19741129 200012 2 005