

**KLASIFIKASI KUALITAS PERGURUAN TINGGI
DI INDONESIA MENGGUNAKAN *RANDOM
FOREST* dan *LOGISTIC REGRESSION***

TESIS

**Oleh:
MASDAR WIYONO
NIM. 220605220007**



**PROGRAM STUDI MAGISTER INFORMATIKA
FEKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**KLASIFIKASI KUALITAS PERGURUAN TINGGI
DI INDONESIA MENGGUNAKAN *RANDOM
FOREST* dan *LOGISTIC REGRESSION***

TESIS

**Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:
MASDAR WIYONO
NIM. 220605220007**

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

**KLASIFIKASI KUALITAS PERGURUAN TINGGI DI
INDONESIA MENGGUNAKAN *RANDOM FOREST*
dan *LOGISTIC REGRESSION***

THESIS

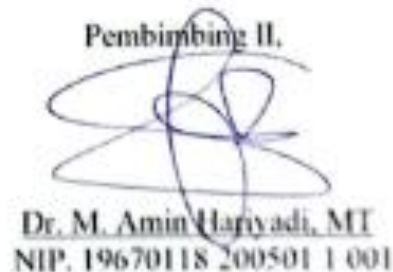
Oleh:
MASDAR WIYONO
NIM. 220605220007

Telah diperiksa dan disetujui untuk diuji:
Tanggal: 29 Oktober 2025

Pembimbing I,


Dr. Gullyon Crysdian, M. Cs
NIP. 19740424 200901 1 008

Pembimbing II,


Dr. M. Amin Haryadi, MT
NIP. 19670118 200501 1 001

Mengetahui,
Ketua Program Studi Magister Informatika
Ilmu Sains dan Teknologi
Universitas Maulana Malik Ibrahim Malang




Prof. Dr. Muhammad Faisal, M. I
NIP. 19740424 200901 1 008

**KLASIFIKASI KUALITAS PERGURUAN TINGGI DI
INDONESIA MENGGUNAKAN *RANDOM FOREST*
dan *LOGISTIC REGRESSION***

THESIS

Oleh:
MASDAR WIYONO
NIM. 220605220007

Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal: 29 Oktober 2025

Susunan Dewan Penguji

Penguji I : Dr. Zainal Abidin, M. Kom
NIP 19760613 200501 1 004

Penguji II : Dr. Agung Teguh Wibowo Almais, M.T
NIP 19860301 202321 1 016

Pembimbing I : Dr. Cahyo Crysdian, M. Cs
NIP. 19740424 200901 1 008

Pembimbing II : Dr. M. Amin Hariyadi, MT
NIP. 19670118 200501 1 001

()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Magister Informatika
Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Prof. Dr. Muhammad Faisal, M.T
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : MASDAR WIYONO
NIM : 220605220007
Fakultas / Prodi : Sains dan Teknologi / Magister Informatika
Fakultas : KLASIFIKASI KUALITAS PERGURUAN TINGGI DI
INDONESIA MENGGUNAKAN *RANDOM FOREST*
dan *LOGISTIC REGRESSION*

Menyatakan dengan sebenarnya bahwa Tesis yang saya tulis ini benar-banar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Tesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 17 November 2025
Yang membuat pernyataan,



Masdar Wiyono
NIM. 220605220007

MOTTO

***“ILMU YANG BERMANFAAT AKAN MEMBAWA KITA KEDALAM
KEBAHAGIAAN SEJATI”***

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Syukur alhamdulillah penulis hanturkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Tesis ini dengan baik.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan jazakumullah ahsanal jaza' kepada semua pihak yang telah membantu terselesaikannya Tesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Dr. Cahyo Crysdiyan, M. Cs dan Bapak Dr. M. Amin Hariyadi, MT selaku dosen pembimbing Tesis, yang telah banyak memberikan pengarahan dan pengalaman yang berharga.
2. Bapak Dr. Zainal Abidin, M. Kom dan Dr. Agung Teguh Wibowo Almais, M.T selaku dosen penguji Tesis, yang telah memberikan kritik dan arahan yang membangun keilmuan penulis menjadi lebih baik khususnya di penulisan ilmiah.
3. Segenap sivitas akademika Program Studi Magister Informatika, terutama seluruh Bapak/ Ibu dosen, terima kasih atas segenap ilmu dan bimbingannya.
4. Bapak/Ibu Rektor, Dekan dan Kaprodi, saya ucapkan terimakasih karena sudah memberikan kesempatan bagi penulis untuk belajar di UIN Malang. Berkat Bapak/Ibu saya dapat menambah pengetahuan yang baik.
5. Ayah dan Ibu tercinta yang senantiasa memberikan doa dan restunya kepada penulis dalam menuntut ilmu.
6. Istri dan Anak-anak penulis yang selalu memberikan dukungan penuh kepada penulis untuk menyelesaikan Tesis ini.

7. Semua pihak yang ikut membantu dalam menyelesaikan Tesis ini baik berupa materil maupun moril.

Penulis menyadari bahwa dalam penyusunan Tesis ini masih terdapat kekurangan dan penulis berharap semoga Tesis ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. *Amin Ya Rabbal Alamin.*

Wassalamu'alaikum Wr. Wb.

Malang, 17 November 2025

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN PENGANTAR	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
ABSTRAK.....	xiii
ABSTRACT	xiv
البحث مستخلص	xv
BAB I.....	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Pernyataan Masalah	10
1.3. Tujuan Penelitian	11
1.4. Ruang Lingkup Penelitian	11
1.5. Manfaat Penelitian	12
BAB II	13
STUDI PUSTAKA	13
2.1. Klasifikasi Kualitas Perguruan Tinggi	13
2.2. Kerangka Teori	20
BAB III	26
PROSEDUR PENELITIAN	26
3.1. Prosedur Penelitian	26
3.1.1. Pengumpulan Data	27
3.1.2. Persiapan Data.....	29
3.1.3. Desain Sistem.....	30
3.1.4. Proses Utama.....	33
3.1.5. Implementasi Sistem	34
3.1.6. Eksperimen.....	35
3.2. Instrumen Penelitian	39
BAB IV	41
KLASIFIKASI KUALITAS PERGURUAN TINGGI DENGAN <i>RANDOM FOREST</i>	41
4.1. Desain Model <i>Random Forest</i>	44
4.2. Implementasi Model <i>Random Forest</i>	46
4.2.1. Pelatihan Model RF-A	49
4.2.2. Pelatihan Model RF-B	53
4.2.3. Pelatihan Model RF-C	56
4.2.4. Pelatihan Model RF-D	58
4.2.5. Pelatihan Model RF-E.....	61

4.3.	Ujicoba Model <i>Random Forest</i>	64
4.4.	Kesimpulan	76
BAB V		79
KLASIFIKASI KUALITAS PERGURUAN TINGGI DENGAN <i>LOGISTIC</i>		
<i>REGRESSION</i>		79
5.1.	Desain Model <i>Logistic Regreesion</i>	82
5.2.	Implementasi Model <i>Logistic Regression</i>	85
5.2.1.	Pelatihan Model LR-A	87
5.2.2.	Pelatihan Model LR-B	90
5.2.3.	Pelatihan Model LR-C	93
5.2.4.	Pelatihan Model LR-D	96
5.2.5.	Pelatihan Model LR-E	99
5.3.	Ujicoba Model <i>Logistic Regression</i>	102
5.4.	Kesimpulan	114
BAB VI.....		118
PEMBAHASAN.....		118
BAB VII.....		136
KESIMPULAN.....		136
7.1.	Kesimpulan	136
7.2.	Keterbatasan.....	137
DAFTAR PUSTAKA.....		138

DAFTAR GAMBAR

Gambar 1. 1 Jumlah perguruan tinggi di Indonesia tahun 2023	1
Gambar 2. 1 Kerangka teori Klasifikasi Kualitas Perguruan Tinggi di Indonesia (PTI)...	21
Gambar 3. 1 Prosedur penelitian	27
Gambar 3. 2 Desain sistem.....	30
Gambar 4. 1 Arsitektur model <i>Random Forest</i> 300 tree.....	45
Gambar 4. 2 Flowchart Model <i>Random Forest</i>	46
Gambar 4. 3 Grafik train dan validation loss model RF-A	50
Gambar 4. 4 Hasil klasifikasi model RF-A terhadap tingkat akreditasi.....	51
Gambar 4. 5 Grafik train dan validation loss model RF-B	53
Gambar 4. 6 Hasil klasifikasi model RF-B terhadap tingkat akreditasi	54
Gambar 4. 7 Grafik train dan validation loss model RF-C	56
Gambar 4. 8 Hasil klasifikasi model RF-C terhadap tingkat akreditasi	57
Gambar 4. 9 Grafik train dan validation loss model RF-D	59
Gambar 4. 10 Hasil klasifikasi model RF-D terhadap tingkat akreditasi.....	60
Gambar 4. 11 Grafik train dan validation loss model RF-E	62
Gambar 4. 12 Hasil klasifikasi model RF-E terhadap tingkat akreditasi	62
Gambar 4. 13 Perbandingan matrik evaluasi model RF-A	65
Gambar 4. 14 Perbandingan matrik evaluasi model RF-B	66
Gambar 4. 15 Perbandingan matrik evaluasi model RF-C	68
Gambar 4. 16 Perbandingan matrik evaluasi model RF-D	69
Gambar 4. 17 Perbandingan matrik evaluasi model RF-E.....	71
Gambar 4. 18 Skor AUC model RF-D	75
Gambar 4. 19 Perbandingan performa model <i>Random Forest</i>	77
Gambar 5. 1 Arsitektur model <i>Logistic Regression</i> pada Klasifikasi Kualitas Perguruan Tinggi.....	83
Gambar 6. 1 Perbandingan akurasi model <i>Random Forest</i> dan <i>Logistic Regression</i>	122
Gambar 6. 2 Performa kinerja model terbaik	133

DAFTAR TABEL

Tabel 2. 1 Kinerja Klasifikasi Kualitas menggunakan <i>Machine Learning</i>	22
Tabel 3. 1 Data Indikator Kualitas Perguruan Tinggi	29
Tabel 3. 2 Data indikator kualitas perguruan tinggi setelah di normalisasi	29
Tabel 3. 3 Data indikator kualitas perguruan tinggi Indonesia	31
Tabel 3. 4 Komposisi data latih dan data uji	36
Tabel 3. 5 Konsep <i>confusion matrix</i>	36
Tabel 3. 6 Interpretasi nilai AUC	39
Tabel 3. 7 Variabel penelitian	39
Tabel 4. 1 Parameter model <i>Random Forest</i>	48
Tabel 4. 2 Distribusi data pelatihan model <i>Random Forest</i>	49
Tabel 4. 3 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-A.....	52
Tabel 4. 4 Performa hasil pengujian model RF-A	53
Tabel 4. 5 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-B.....	55
Tabel 4. 6 Performa hasil pengujian model RF-B	55
Tabel 4. 7 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-C.....	58
Tabel 4. 8 Performa hasil pengujian model RF-C	58
Tabel 4. 9 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-D.....	61
Tabel 4. 10 Performa hasil pengujian model RF-D	61
Tabel 4. 11 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-E.....	63
Tabel 4. 12 Performa hasil pengujian model RF-E	64
Tabel 4. 13 Hasil ujicoba klasifikasi kualitas model <i>Random Forest</i>	64
Tabel 4. 14 Performa model <i>Random Forest</i>	71
Tabel 4. 15 Nilai TP, TN, FP, FN model RF-D.....	73
Tabel 5. 1 Parameter model <i>Logistic Regression</i>	86
Tabel 5. 2 Distribusi pelatihan model <i>Logistic Regression</i>	87
Tabel 5. 3 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-A	90
Tabel 5. 4 Performa hasil pengujian model LR-A	90
Tabel 5. 5 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-B.....	93
Tabel 5. 6 Performa hasil pengujian model LR-B	93
Tabel 5. 7 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-C.....	96
Tabel 5. 8 Performa hasil pengujian model LR-C	96
Tabel 5. 9 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-D	99
Tabel 5. 10 Performa hasil pengujian model LR-D	99
Tabel 5. 11 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-E.....	102
Tabel 5. 12 Performa hasil pengujian model LR-E	102
Tabel 5. 13 Hasil ujicoba klasifikasi kualitas model <i>Logistic Regression</i>	103
Tabel 5. 14 Performa model <i>Logistic Regression</i>	110
Tabel 5. 15 Nilai TP, TN, FP, FN model LR-C.....	111
Tabel 6. 1 Performa kinerja model terbaik (RF-D & LR-C)	123
Tabel 6. 2 Nilai TP TP, TN, FP, FN model terbaik (RF-D & LR-C)	124
Tabel 6. 3 Skor AUC pada performa model terbaik (RF-D & LR-C).....	126

ABSTRAK

Wiyono, Masdar. 2025. **Klasifikasi Kualitas Perguruan Tinggi di Indonesia Menggunakan *Random Forest* dan *Logistic Regression***. Tesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crysdian, M. Cs (II) Dr. M. Amin Hariyadi, MT.

Kata kunci: Klasifikasi Kualitas Perguruan Tinggi, *Random Forest*, *Logistic Regression*.

Penelitian ini bertujuan untuk melakukan Klasifikasi Kualitas Perguruan Tinggi di Indonesia menggunakan Model *Random Forest* dan *Logistic Regression*. Data yang digunakan merupakan data sekunder yang telah diproses dan dibagi dalam beberapa rasio pelatihan dan pengujian. Hasil penelitian menunjukkan bahwa kedua Model memiliki performa yang kompetitif dengan keunggulan masing-masing. Model *Logistic Regression* terbaik (LR-C, rasio 70:30) mencapai akurasi 77,78% dengan nilai AUC tinggi di semua kelas mutu, sehingga cocok untuk analisis yang memerlukan interpretasi yang jelas. Sementara itu, Model *Random Forest* terbaik (RF-D, rasio 60:40) menunjukkan akurasi lebih tinggi, yaitu 81,25%, serta kemampuan yang baik dalam mengenali seluruh kategori mutu. Temuan ini menunjukkan bahwa kombinasi kedua Model dapat digunakan untuk meningkatkan ketepatan klasifikasi, dimana *Logistic Regression* efektif sebagai model dasar yang mudah dipahami, sedangkan *Random Forest* berperan sebagai model pendukung untuk validasi dan penguatan hasil. Penelitian ini masih memiliki keterbatasan, seperti penggunaan indikator administratif saja, distribusi kelas yang tidak seimbang, serta model yang belum adaptif terhadap perubahan mutu secara dinamis. Penelitian selanjutnya perlu melibatkan indikator yang lebih luas agar klasifikasi kualitas perguruan tinggi dapat dilakukan secara lebih komprehensif.

ABSTRACT

Wiyono, Masdar. 2025. **Classify the Quality of Higher Education Institutions in Indonesia Using the Random Forest and Logistic Regression Methods**. Thesis. Master of Informatics Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang. Supervisor: (I) Dr. Cahyo Crysdian, M. Cs (II) Dr. M. Amin Hariyadi, MT.

Keywords: Higher Education Quality Classification, Random Forest, Logistic Regression.

This study aims to classify the quality of higher education institutions in Indonesia using the Random Forest and Logistic Regression methods. The research utilizes secondary data that has been pre-processed and split into several training-testing ratios. The results show that both methods demonstrate competitive performance with their respective strengths. The best Logistic Regression model (LR-C, 70:30 ratio) achieved an accuracy of 77.78% with high AUC values across all quality categories, making it suitable for analyses requiring clear interpretability. Meanwhile, the best Random Forest model (RF-D, 60:40 ratio) achieved a higher accuracy of 81.25% and showed strong capability in identifying all quality categories. These findings indicate that combining both methods can enhance classification accuracy, where Logistic Regression serves effectively as an interpretable baseline model, while Random Forest provides additional support for validation and strengthening of results. This study has several limitations, including the use of administrative indicators only, imbalanced class distribution, and models that are not adaptive to dynamic quality changes. Future research should incorporate broader indicators to achieve a more comprehensive classification of higher education quality.

البحث مستخلص

ويونو، مسدار. 2025. تصنيف جودة مؤسسات التعليم العالي في إندونيسيا باستخدام خوارزميتي الغابة العشوائية والانحدار اللوجستي. رسالة ماجستير، برنامج الماجستير في علوم المعلوماتية، كلية العلوم والتكنولوجيا، جامعة إسلام نيجري مولانا مالك إبراهيم مالانج. المشرفان: (1) د. كاهيو كريسديان، ماجستير علوم، (2) د. م. أمين هاريادي، ماجستير هندسة.

الكلمات المفتاحية: تصنيف جودة مؤسسات التعليم العالي، الغابة العشوائية، الانحدار اللوجستي.

يهدف هذا البحث إلى تصنيف جودة مؤسسات التعليم العالي في إندونيسيا باستخدام نموذجي الغابة العشوائية و الانحدار اللوجستي. تعتمد البيانات المستخدمة على بيانات ثانوية تمت معالجتها وتقسيمها إلى عدة نسب للتدريب والاختبار. وتُظهر نتائج البحث أن النموذجين يمتلكان أداءً تنافسياً مع ميزات مميزة لكل منهما. إذ حقق أفضل نموذج للانحدار اللوجستي (LR-C) بنسبة (70:30) دقة بلغت 77.78% مع قيمة AUC مرتفعة في جميع فئات الجودة، مما يجعله مناسباً للتحليلات التي تتطلب تفسيراً واضحاً. بينما أظهر أفضل نموذج للغابة العشوائية (RF-D) بنسبة (60:40) دقة أعلى بلغت 81.25%، إضافةً إلى قدرته الجيدة على التعرف على جميع فئات الجودة. تشير هذه النتائج إلى أن الجمع بين النموذجين يمكن أن يُستخدم لتحسين دقة التصنيف، حيث يُعد الانحدار اللوجستي نموذجاً أساسياً فعالاً وسهل الفهم، بينما تعمل الغابة العشوائية كنموذج داعم للتأكيد وتعزيز النتائج. ومع ذلك، ما يزال البحث يعاني من بعض القيود، مثل الاعتماد على مؤشرات إدارية فقط، وعدم توازن توزيع الفئات، فضلاً عن عدم قدرة النموذج على التكيف مع التغيرات الديناميكية في الجودة. وينبغي أن تتضمن الأبحاث المستقبلية مؤشرات أكثر شمولاً من أجل الحصول على تصنيف أكثر تكاملاً لجودة مؤسسات التعليم العالي.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perguruan tinggi merupakan jenjang pendidikan lanjutan setelah pendidikan menengah yang bertujuan untuk membekali peserta didik agar dapat menjadi bagian dari masyarakat dengan kompetensi akademik dan professional. Lulusan perguruan tinggi diharapkan mampu menerapkan, mengembangkan, serta menciptakan ilmu pengetahuan, teknologi, dan seni (Karim, 2020).

Saat ini, perkembangan perguruan tinggi di Indonesia terus mengalami peningkatan (Nulhaqim *et al.*, 2016). Hal ini terlihat dari bertambahnya jumlah perguruan tinggi baru setiap tahunnya, baik dalam bentuk universitas, akademi, institut, politeknik, maupun sekolah tinggi, yang berstatus negeri maupun swasta, seperti yang ditunjukkan pada Gambar 1.1. Berdasarkan data PDDIKTI tahun 2023, terdapat total 4.437 Perguruan Tinggi di Indonesia (RISET, 2022).



Gambar 1. 1 Jumlah perguruan tinggi di Indonesia tahun 2023

Perguruan tinggi di Indonesia memiliki peran yang signifikan dalam mencerdaskan kehidupan bangsa. Berdasarkan Undang-Undang (UU) Nomor 12 Tahun 2012, perguruan tinggi didefinisikan sebagai lembaga pendidikan yang menyelenggarakan pendidikan tinggi (Indonesia, 2012). Di Indonesia, pendidikan tinggi terbagi menjadi dua jenis, yaitu perguruan tinggi negeri (PTN) yang didirikan dan dikelola oleh pemerintah, serta perguruan tinggi swasta (PTS) yang diselenggarakan oleh masyarakat melalui yayasan atau badan penyelenggara lainnya. Mengacu pada Undang-Undang Nomor 20 Tahun 2003 dan Peraturan Pemerintah (PP) Nomor 60 Tahun 1999 tentang Pendidikan Tinggi, perguruan tinggi, baik melalui fakultas maupun program studinya, memiliki tugas utama dalam mengembangkan dan meningkatkan kualitas sumber daya manusia guna mendukung kemajuan bangsa serta peradaban (Indonesia, 2006).

Peningkatan jumlah perguruan tinggi di Indonesia, khususnya Perguruan Tinggi Negeri (PTN), menyebabkan persaingan yang semakin kompetitif dalam hal kualitas pendidikan di Indonesia. Untuk memastikan mutu pendidikan yang optimal, pemerintah telah menetapkan standar nasional pendidikan tinggi yang mencakup aspek pendidikan, penelitian, dan pengabdian kepada masyarakat. Namun, tantangan muncul dalam implementasi dan pengukuran standar tersebut, terutama dalam memastikan setiap perguruan tinggi memenuhi kriteria yang telah ditetapkan. Kurangnya sistem evaluasi yang efektif dan mekanisme pengawasan yang ketat dapat berdampak pada disparitas kualitas antar perguruan tinggi. Selain itu, klasifikasi pengukuran kualitas perguruan tinggi di Indonesia masih banyak bergantung pada metode evaluasi administratif dan proses akreditasi

konvensional, yang meskipun bermanfaat, sering kali belum sepenuhnya mencerminkan performa nyata sebuah institusi. Evaluasi semacam ini cenderung bersifat periodik, statis, dan kurang fleksibel dalam menangkap dinamika mutu pendidikan secara real time.

Oleh karena itu, diperlukan metode pengukuran yang objektif, transparan, dan berkelanjutan untuk menilai kualitas perguruan tinggi di Indonesia. Penelitian-penelitian sebelumnya umumnya hanya menggunakan satu algoritma atau menilai satu aspek kualitas saja, seperti akreditasi atau persepsi lulusan. Hingga kini, masih sedikit studi yang secara langsung membandingkan kinerja model Random Forest dan Logistic Regression dalam menganalisis kualitas perguruan tinggi secara lebih menyeluruh berbasis data aktual.

Urgensi dan kebaruan dari penelitian ini terletak pada penerapan metode data mining berbasis *Random Forest* dan *Logistic Regression* secara spesifik untuk konteks perguruan tinggi di Indonesia, yang hingga kini masih minim eksplorasi mendalam, terutama dalam skala nasional dan berbasis data indikator kualitas perguruan tinggi. Meskipun pendekatan serupa telah digunakan dalam studi internasional untuk pemeringkatan global seperti QS-WUR, belum banyak riset yang mengkaji penggunaan dua algoritma ini secara komparatif dalam menilai kualitas perguruan tinggi negeri di Indonesia.

Syahrir *et al.*, (2023) menjelaskan bahwa untuk meningkatkan dan menilai kualitas Perguruan Tinggi Negeri (PTN), pemerintah menerapkan Indikator Kinerja Utama (IKU) sebagai standar evaluasi dalam berbagai aspek. IKU

mencakup beberapa indikator, seperti tingkat keberhasilan lulusan dalam memperoleh pekerjaan yang layak, kualitas dosen serta tenaga kependidikan, jumlah penelitian dan publikasi yang memberikan dampak, serta keterlibatan perguruan tinggi dalam kerja sama dengan industri dan masyarakat. IKU berfungsi sebagai alat ukur utama dalam menilai kinerja PTN, sekaligus memastikan bahwa perguruan tinggi dapat memberikan kontribusi nyata bagi pembangunan nasional (Syahrir *et al.*, 2023).

Dalam menilai kualitas sebuah perguruan tinggi, kita tidak bisa hanya mengandalkan satu sisi saja. Diperlukan berbagai indikator yang bisa memberi gambaran utuh tentang seberapa baik sebuah institusi menjalankan fungsinya. Salah satu indikator awal yang bisa dilihat adalah jumlah program studi yang dimiliki. Hal ini mencerminkan seberapa luas cakupan bidang keilmuan yang ditawarkan kampus tersebut. Semakin beragam program studi yang ada, maka semakin besar peluang mahasiswa untuk mengembangkan potensi diri sesuai minat dan kebutuhan zaman (Singgih, 2008). Selain itu, jumlah mahasiswa baru yang diterima setiap tahun juga bisa menjadi tanda bahwa masyarakat memiliki kepercayaan tinggi terhadap institusi tersebut (Angkotasari & Watianan, 2021).

Indikator jumlah mahasiswa aktif mencerminkan kapasitas kampus dan kemampuan perguruan tinggi dalam menjaga keberlangsungan proses belajar-mengajar (Andriani, 2012). Di sisi lain, jumlah lulusan menunjukkan keberhasilan perguruan tinggi dalam menuntaskan pendidikan mahasiswanya serta kontribusinya dalam menyediakan tenaga terdidik bagi masyarakat dan negara (Asmawi, 2005).

Peran penting juga dimainkan oleh dosen, sebagai ujung tombak dalam pembelajaran dan riset. Jumlah dosen yang memadai, serta seimbang dengan jumlah mahasiswa, menjadi kunci bagi berlangsungnya proses akademik yang efektif (Fitrah, 2018). Di samping itu, kegiatan riset juga tak kalah penting. Lewat Skor SINTA, baik dalam tiga tahun terakhir maupun secara keseluruhan, kita bisa melihat sejauh mana kampus berkontribusi dalam bidang penelitian dan publikasi ilmiah (Nasution, 2016). Skor ini membantu menilai kualitas dan konsistensi institusi dalam menyumbang ilmu pengetahuan baru.

Terakhir, kita juga perlu melihat pada aspek akreditasi yang merupakan bentuk pengakuan resmi terhadap kualitas sebuah perguruan tinggi. Akreditasi menunjukkan bahwa sebuah institusi telah dinilai berdasarkan berbagai standar penting seperti mutu pendidikan, pengelolaan kampus, hingga kontribusi terhadap masyarakat (Mitra, 2023). Semua indikator ini, jika dianalisis menggunakan pendekatan berbasis data seperti *Random Forest* dan *Logistic Regression*, dapat memberikan gambaran yang lebih objektif tentang mutu pendidikan tinggi. Harapannya, pendekatan ini bisa membantu menghasilkan kebijakan yang lebih tepat dalam meningkatkan kualitas kampus-kampus di Indonesia.

Proses pengukuran kualitas dapat dilihat di dalam Alqur'an dan dapat dikaitkan dengan prinsip penilaian secara objektif, keseimbangan, dan keadilan termasuk dalam bidang pekerjaan dan pendidikan salah satunya yang terdapat dalam surat Al-Mulk 67:2. Kemudian di tafsirkan oleh Ibn Katsir pada kitab karyanya yang berjudul *Tafsir al-Qur'an al-'Azhim*, berikut ayat dan tafsirnya :

الَّذِي خَلَقَ الْمَوْتَ وَالْحَيَاةَ لِيَبْلُوَكُمْ أَيُّكُمْ أَحْسَنُ عَمَلًا وَهُوَ الْعَزِيزُ الْغَفُورُ

“yaitu yang menciptakan kematian dan kehidupan untuk menguji kamu, siapa di antara kamu yang lebih baik amalnya. Dia Maha Perkasa lagi Maha Pengampun” (QS. Al-Mulk 67:2, Kemenag, 2019).

Ayat Al-Mulk 67:2 menjelaskan bahwa Allah menciptakan hidup dan mati sebagai ujian untuk melihat kualitas amal manusia. Quraish Shihab (2002) dalam Tafsir al-Mishbah menekankan bahwa ayat ini mengajarkan pentingnya kualitas dan kesungguhan dalam setiap aktivitas manusia. Prinsip ini dapat dijadikan dasar etik dalam pengukuran kualitas perguruan tinggi, yaitu bahwa keberhasilan suatu institusi tidak hanya bergantung pada kuantitas sumber daya, tetapi pada kualitas kontribusi dan manfaat yang dihasilkan bagi masyarakat dan perkembangan ilmu pengetahuan (Syafii *et al.*, 2023). Surat Al-Hujurāt ayat 13 yaitu:

يَا أَيُّهَا النَّاسُ إِنَّا خَلَقْنَاكُمْ مِنْ ذَكَرٍ وَأُنْثَىٰ وَجَعَلْنَاكُمْ شُعُوبًا وَقَبَائِلَ لِتَعَارَفُوا إِنَّ أَكْرَمَكُمْ عِنْدَ اللَّهِ أَتْقَاكُمْ إِنَّ اللَّهَ عَلِيمٌ خَبِيرٌ

“Wahai manusia, sesungguhnya Kami telah menciptakan kamu dari seorang laki-laki dan perempuan. Kemudian, Kami menjadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah adalah orang yang paling bertakwa. Sesungguhnya Allah Maha Mengetahui lagi Maha Teliti”.(QS. Al-Hujurāt (49):13, Kemenag, 2019)

Menurut Quraish Shihab (2002) dalam *Tafsir Al-Mishbah*, QS. Al-Hujurāt ayat 13 menegaskan prinsip kesetaraan manusia dengan menyatakan bahwa seluruh manusia berasal dari satu laki-laki dan satu perempuan, sehingga tidak ada alasan untuk merasa lebih tinggi berdasarkan perbedaan etnis, suku, atau kelompok sosial. Perbedaan manusia ke dalam bangsa-bangsa dan suku-suku merupakan kehendak Allah agar mereka saling mengenal, bekerja sama, dan

membangun relasi sosial yang harmonis. Quraish Shihab menjelaskan bahwa kemuliaan seseorang tidak ditentukan oleh identitas lahiriah, tetapi oleh ketakwaannya yang tercermin melalui kesadaran moral dan kontribusi positif kepada masyarakat. Dengan demikian, ayat ini mempertegas bahwa keberagaman adalah sarana untuk membangun kehidupan sosial yang lebih baik, sementara kualitas seseorang diukur dari nilai-nilai ketakwaan dan akhlaknya, bukan dari aspek-aspek sosial yang bersifat lahiriah (Quraish Shihab, 2002).

Berdasarkan penafsiran tersebut apabila dikaitkan dengan konteks pengukuran dan pengklasifikasian kualitas perguruan tinggi harus didasarkan pada sejumlah parameter yang relevan agar proses perbandingan dan analisis dapat dilakukan secara lebih objektif, menyeluruh, dan terukur. Prinsip “kualitas” dalam QS. Al-Hujurāt 49:13 yang menekankan nilai ketakwaan sebagai indikator kemuliaan secara metodologis dapat diarahkan pada pentingnya indikator-indikator yang mencerminkan kontribusi, kinerja, serta dampak nyata suatu institusi pendidikan. Sejalan dengan itu, penelitian ini menggunakan beberapa atribut sebagai dasar penilaian. Sehingga analisis kualitas dapat dilakukan secara komprehensif dan sesuai dengan prinsip objektivitas dalam evaluasi institusi.

Selanjutnya, metode yang dapat digunakan untuk mengukur kualitas perguruan tinggi di Indonesia cenderung banyak. Salah satunya adalah data mining untuk melakukan klasifikasi pengukuran kualitas secara mendalam. Data mining memungkinkan analisis berbagai aspek yang memengaruhi kualitas perguruan tinggi, seperti kinerja akademik, tingkat kepuasan mahasiswa, daya

serap lulusan di dunia kerja, serta produktivitas penelitian dan publikasi ilmiah (Romero & Ventura, 2007). Dengan teknik seperti klasifikasi, klusterisasi, dan asosiasi, data mining dapat mengungkap pola dan tren yang tidak terlihat secara langsung dalam data (Rohmawati *et al.*, 2015).

Beberapa model yang dapat digunakan untuk mengukur kualitas seperti *Linear Regression*, *Decision Tree*, *Support Vector Machine (SVM)*, *Random Forest*, dan *Logistic Regression*. Salah satu penerapan model klasifikasi pengukuran kualitas perguruan tinggi berdasarkan penelitian sebelumnya yang dilakukan oleh Estrada-Real & Cantu-Ortiz, (2022) menggunakan Algoritma *Logistic Regression*, *Support Vector Machine (SVM)*, dan *Random Forest* untuk mengukur kinerja universitas dalam peringkat QS World University Rankings (QS-WUR). Penelitian ini menggunakan data selama sepuluh tahun untuk memprediksi kinerja universitas, mengklasifikasikan institusi ke dalam kelompok Top 100 dan Top 200. Hasilnya menunjukkan bahwa di antara algoritma yang digunakan tersebut, algoritma *Logistic Regression* dan *Random Forest* menunjukkan akurasi tertinggi dan punya kesamaan angka yaitu 0.9948 atau 99.48%.

Konsep model *Random Forest* didasarkan pada prinsip *Ensemble Learning*, yaitu menggabungkan banyak pohon keputusan untuk menghasilkan klasifikasi yang lebih akurat dan stabil. Model ini terbukti efektif dalam menangani data berukuran besar dan kompleks dengan berbagai jenis fitur. Kelebihan *Random Forest* terletak pada kemampuannya menangani *outlier*, data yang tidak seimbang,

serta mengurangi risiko overfitting. Sementara itu, kekurangannya adalah waktu pelatihan yang relatif lebih lama dan kurangnya interpretasi yang sederhana dibandingkan model linier (Breiman, 2001).

Sementara itu, *Logistic Regression* merupakan model klasifikasi yang memprediksi probabilitas suatu data masuk ke dalam kategori tertentu berdasarkan fungsi logistik. Model ini lebih mudah diinterpretasikan karena menghasilkan koefisien yang menunjukkan pengaruh masing-masing fitur terhadap hasil klasifikasi. Namun, kekurangan terletak pada keterbatasannya dalam menangani hubungan non-linear serta performa yang kurang optimal jika data sangat kompleks atau memiliki banyak fitur interaksi (Hosmer *et al.*, 2013).

Penelitian ini berfokus pada penggunaan algoritma *Random Forest* dan *Logistic Regression* untuk menilai kualitas perguruan tinggi negeri (PTN) di Indonesia. Pendekatan ini tidak hanya membantu mengukur kinerja perguruan tinggi secara lebih objektif, tetapi juga bertujuan memberikan gambaran yang lebih menyeluruh dan adil terhadap kontribusi mereka dalam dunia pendidikan dan masyarakat.

Keunikan dan nilai penting dalam penelitian ini adalah penerapan dua algoritma data mining secara langsung dalam konteks penilaian kualitas perguruan tinggi di Indonesia. Hal tersebut merupakan sesuatu yang hingga kini masih jarang dilakukan, apalagi dalam lingkup nasional dan berdasarkan data indikator kualitas perguruan tinggi. Selama ini, penilaian mutu perguruan tinggi lebih sering dilakukan melalui mekanisme konvensional, seperti akreditasi atau

laporan administratif. Terkadang yang dilakukan belum sepenuhnya menggambarkan kondisi yang sebenarnya di lapangan.

Dengan memanfaatkan teknologi yang semakin berkembang, penelitian ini ingin menawarkan cara baru yang lebih terbuka, akurat, dan bisa dipertanggungjawabkan dalam menilai kualitas pendidikan tinggi. Harapannya, hasil dari studi ini bisa menjadi pijakan dalam merancang kebijakan pendidikan yang lebih tepat sasaran, sekaligus mendorong perguruan tinggi untuk terus berinovasi dan berkontribusi nyata bagi kemajuan bangsa.

Penelitian ini mengeksplorasi terkait dengan kinerja algoritma data mining menggunakan model *Random Forest* dan *Logistic Regression* untuk mengukur kualitas perguruan tinggi di Indonesia, khususnya PTN secara efektif dan optimal untuk mengetahui tingkat performa atau kinerja dari model tersebut. Penelitian ini diharapkan dapat memberikan rekomendasi strategis bagi pemangku kepentingan dalam meningkatkan mutu pendidikan tinggi di Indonesia. Selain itu, dengan memahami berbagai aspek yang terkait dengan klasifikasi pengukuran kualitas, diharapkan hasil penelitian ini dapat berkontribusi dalam perumusan kebijakan pendidikan tinggi yang lebih efektif dan berkelanjutan.

1.2. Pernyataan Masalah

1. Bagaimana membangun model klasifikasi untuk mengukur kualitas perguruan tinggi di Indonesia menggunakan *Random Forest* dan *Logistic Regression* ?

2. Bagaimana performa kedua model tersebut dalam mengklasifikasikan kualitas perguruan tinggi di Indonesia berdasarkan metrik evaluasi yang digunakan ?

1.3. Tujuan Penelitian

1. Mengembangkan model klasifikasi untuk mengukur kualitas perguruan tinggi di Indonesia menggunakan metode Random Forest dan Logistic Regression.
2. Mengevaluasi dan membandingkan performa kedua model tersebut dalam mengklasifikasikan kualitas perguruan tinggi di Indonesia berdasarkan metrik evaluasi yang relevan.

1.4. Ruang Lingkup Penelitian

1. Data yang digunakan pada penelitian ini adalah data Statistik Pendidikan Tinggi tahun 2023 yang terbitkan oleh Sekretariat Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi dan data skor publikasi penelitian tiga tahun terakhir dan keseluruhan.
2. Data di peroleh dari website Pangkalan Data Pendidikan Tinggi (PDDIKTI) dan website SINTA pada bulan 26 April 2025.
3. Atribut yang digunakan pada penelitian ini Perguruan Tinggi, Skor_IKU, Program_Studi, Mahasiswa_Baru, Mahasiswa_Terdaftar, Lulusan, Dosen, Skor_Sinta_3Yr, Skor_Sinta_Overall, dan Akreditasi.
4. Variabel keluaran yang diharapkan adalah kualitas Perguruan Tinggi di Indonesia.

1.5. Manfaat Penelitian

1. Bagi pihak pemangku kepentingan di perguruan tinggi Indonesia dapat dijadikan sebagai rujukan dalam merancang strategi untuk meningkatkan mutu pendidikan di perguruan tinggi Indonesia.
2. Dapat dijadikan sebagai bahan pertimbangan dalam perumusan kebijakan pendidikan tinggi secara efektif dan berkesinambungan.
3. Dapat dijadikan sebagai bahan referensi bagi para peneliti yang akan melakukan penelitian sejenis.

BAB II

STUDI PUSTAKA

2.1. Klasifikasi Kualitas Perguruan Tinggi

Pada studi yang dilakukan oleh Claassen, (2015) menganalisis data dari delapan sistem pemeringkatan universitas nasional dan internasional. Tujuan dari penelitian ini yaitu untuk mengukur kualitas universitas menggunakan model *Bayesian Hierarkis*. Akurasi yang di hasilkan dalam penelitian ini tidak di jelaskan secara spesifik, namun peneliti menyebutkan bahwa penelitian ini memiliki interval kredibel akurasi nilai sekitar 95% yang berarti akurasi nilai kualitas universitas berada pada rentang tersebut berdasarkan nilai akurasi yang merupakan hasil dari ukuran statistik yang menggambarkan seberapa baik model dapat menjelaskan variasi dalam menilai universitas. Hal ini menunjukkan bahwa perkiraan tersebut tidak pasti. Studi ini memberikan estimasi yang lebih transparan, meskipun ada keterbatasan dari sistem yang tidak disertakan. Pada penelitian selanjutnya, penulis menyarankan untuk mengurangi bias dan meningkatkan model evaluasi.

Pada penelitian yang dilakukan oleh Zhang *et al.*, (2021) membahas tentang evaluasi kualitas pengajaran dengan menggabungkan algoritma genetika dengan Artificial Neural Network (ANN) untuk meningkatkan akurasi dan efisiensi dalam menilai kualitas pengajaran di perguruan tinggi. Penelitian ini bertujuan untuk mengembangkan metode evaluasi yang lebih ilmiah, akurat, dan cepat dibandingkan metode tradisional. Metode yang digunakan melibatkan optimasi bobot jaringan saraf menggunakan algoritma genetika, yang membantu mengatasi

kelemahan *Backpropagation (BP)* tradisional, seperti konvergensi lambat dan risiko terjebak dalam minimum lokal. Berdasarkan nilai rata-rata hasil eksperimen dari metode yang digunakan, hasil menunjukkan metode yang diusulkan menghasilkan akurasi lebih tinggi dengan nilai sekitar 95.60% dari pada BP tradisional yang memiliki rata-rata akurasi sekitar 94.16%.

Selanjutnya, Mustakim & Kamal, (2021) melakukan pengukuran kualitas perguruan tinggi di Indonesia dengan cara mengklasifikasikan manajemen kualitas pendidikan berdasarkan parameter standar kualitasnya. Penelitian ini menggunakan model K-Means untuk mengelompokkan data berdasarkan Kesamaan Karakteristik. Hasil penelitian ini menunjukkan bahwa model K-Means dapat mengklasifikasikan tingkat kualitas dengan memperoleh nilai akurasi sekitar 95.84%. Jawa Barat dan Jawa Timur memiliki kemiripan karakteristik dan tergabung dalam kluster yang sama. Sekitar 5,88% provinsi memenuhi syarat untuk menjadi model percontohan dalam manajemen standar pendidikan yang berkualitas. Kelemahan dari penelitian ini yaitu terbatas oleh data sekunder yang tersedia, yang bisa mempengaruhi kualitas analisis.

Penelitian oleh Estrada-Real & Cantu-Ortiz, (2022) memperkenalkan model pengukuran kualitas perguruan tinggi yang menggunakan algoritma statistik dan *Machine Learning* untuk mengukur kinerja universitas dalam peringkat *QS World University Rankings (QS-WUR)*. Metodologi yang dikembangkan dalam penelitian ini bertujuan untuk memprediksi Skor Akhir mitra universitas berdasarkan kinerja historis, dengan tingkat kesalahan hanya berkisar satu hingga dua poin dari total 100. Studi ini menerapkan model klasifikasi *Machine Learning*

dengan algoritma *Logistic Regression*, *Support Vector Machine (SVM)*, dan *Random Forest* untuk menilai kualitas universitas berdasarkan peringkat QS-WUR. Terdapat dua metode yang menghasilkan akurasi tertinggi di antara metode yang digunakan, yaitu *Random Forest* dan *Logistic Regression*, yang keduanya mencapai tingkat akurasi yang sama yaitu sebesar 0,9948 atau 99,48%. Penelitian ini menggunakan data selama sepuluh tahun untuk memprediksi kinerja universitas, mengklasifikasikan institusi ke dalam kelompok Top 100 dan Top 200. Pendekatan ini membantu universitas dalam mengidentifikasi indikator utama yang dapat meningkatkan peringkat dan kualitas mereka secara keseluruhan. Meskipun penerapan berbagai algoritma pada *Machine Learning* telah berhasil, penelitian ini mengakui adanya keterbatasan dalam penggunaan data historis. Model prediksi ini mungkin kurang mampu menangkap perubahan mendadak dalam dinamika universitas atau faktor eksternal yang dapat memengaruhi peringkat, seperti perubahan tren pendidikan global atau ketersediaan pendanaan. Keterbatasan ini dapat berdampak pada ketahanan model dalam memprediksi peringkat universitas di masa depan. Sebagai langkah penelitian selanjutnya, model prediksi yang dikembangkan dalam studi ini dapat diterapkan pada dataset peringkat universitas lainnya, misalnya menggunakan data Indikator Kinerja Utama (IKU) pada perguruan tinggi. Dengan demikian, penelitian ini dapat memberikan pemahaman yang lebih luas mengenai berbagai faktor yang memengaruhi peringkat universitas. Hal ini juga dapat membantu universitas dalam mengambil langkah strategis untuk meningkatkan kinerja dan reputasi akademik mereka dalam jangka panjang.

Penelitian yang dilakukan oleh Ibañez Ramírez *et al.*, (2022) bertujuan untuk menilai kualitas universitas negeri di Kolombia tengah berdasarkan cara hidup mahasiswa. Studi ini menggunakan algoritma klasifikasi *Decision Tree J48* dari platform Weka untuk mengukur kualitas uiversitas berdasarkan kehidupan mahasiswa. Analisis dilakukan berdasarkan 10 variabel independen, seperti beban akademik, sumber daya ekonomi, dan kondisi emosional, yang diperoleh dari 127 survei mahasiswa. Hasil penelitian menunjukkan bahwa faktor yang paling berpengaruh terhadap kualitas universitas berdasarkan kehidupan mahasiswa adalah lingkungan kampus, kondisi tempat tinggal, keadaan emosional, dan hubungan dengan dosen. Model klasifikasi yang digunakan mencapai tingkat efektivitas dengan nilai akurasi sekitar 80%. Temuan ini menyoroti bagaimana faktor eksternal memainkan peran penting dalam pengalaman mahasiswa di universitas serta implikasinya terhadap pengembangan kebijakan kampus yang lebih baik. Sebagai langkah penelitian selanjutnya, dapat dilakukan eksplorasi lebih lanjut terhadap dampak variabel eksternal lainnya, seperti sistem dukungan sosial, akses terhadap layanan kesehatan mental, dan keterlibatan dalam komunitas, guna memperoleh pemahaman yang lebih menyeluruh mengenai faktor-faktor yang memengaruhi kesejahteraan mahasiswa terhadap kualitas universitas tersebut.

Pada studi yang dilakukan oleh Zhong, (2023) membahas tentang evaluasi dan analisis kualitas pengajaran dosen universitas dengan menggunakan model klasifikasi *Machine Learning*. Penelitian ini membandingkan algoritma *Support Vector Machine* (SVM) dengan modifikasi GA-SVM. Model GA-SVM berasal

dari Algoritma Genetika (GA) yang mengoptimalkan parameter SVM, sehingga meningkatkan akurasi dan kinerja dari model tersebut. Hasil eksperimen menunjukkan bahwa metode GA-SVM mengurangi waktu pelatihan dan pengujian sekaligus memberikan skor evaluasi yang lebih akurat dibandingkan metode SVM tradisional. Rata-rata akurasi GA-SVM sekitar 98,36%, lebih tinggi 11,64% dibandingkan SVM dengan nilai akurasi sekitar 86,72%. Pendekatan ini menunjukkan kegunaan *Machine Learning* dalam menilai kualitas pengajaran di universitas. Berdasarkan hasil penelitian ini, penulis berharap metode evaluasi berbasis algoritma GA-SVM dapat digunakan secara luas di sektor pendidikan untuk meningkatkan kualitas pendidikan dan pengajaran.

Selanjutnya, penelitian yang dilakukan oleh Li, (2023) menyoroti keterbatasan metode konvensional dalam Evaluasi Kualitas Pengajaran (*Teaching Quality Evaluation/TQE*) dan mengusulkan model baru yang mengintegrasikan algoritma *Particle Swarm Optimization (PSO)* yang telah ditingkatkan dengan *Backpropagation Neural Network (BPNN)* untuk meningkatkan proses evaluasi di institusi pendidikan tinggi. Model ini dirancang untuk mengoptimalkan bobot koneksi awal dan ambang batas, sehingga meningkatkan kemampuan optimasi global, kecepatan konvergensi, dan akurasi evaluasi. Hasil penelitian menunjukkan bahwa algoritma PSO-BPNN yang telah ditingkatkan secara signifikan mengurangi kesalahan prediksi dibandingkan dengan algoritma BP tradisional, dengan selisih kesalahan berkisar antara 0,012 atau 1,2% hingga 0,0144 atau 1,44% pada lima sampel yang diuji. Jika di rata-rata maka nilai errornya sekitar 1,32%. Hal ini membuktikan efektivitas model dalam

meningkatkan akurasi evaluasi kualitas pengajaran serta memberikan wawasan yang berharga bagi pengembangan praktik TQE di perguruan tinggi. Melalui pendekatan ini, penelitian ini mengatasi keterbatasan metode tradisional yang sering kali tidak menghasilkan evaluasi yang optimal. Dengan meningkatkan kemampuan optimasi global, kecepatan konvergensi, dan akurasi model TQE, pendekatan berbasis PSO-BPNN ini dapat menjadi solusi yang lebih andal dalam menilai kualitas pengajaran dan meningkatkan standar pendidikan tinggi.

Sedangkan pada studi yang dilakukan oleh Haerani *et al.*, (2023) melakukan pengukuran kualitas perguruan tinggi menggunakan model klasifikasi *Machine Learning*, khususnya Metode *Decision Tree* (C4.5), untuk menganalisis data akademik dan mengukur kualitas universitas dengan menilai tingkat keberhasilan mahasiswa. Model ini membantu universitas dalam mengambil keputusan yang lebih tepat terkait kinerja mahasiswa dan tingkat retensi mereka, sehingga berkontribusi pada peningkatan kualitas pendidikan tinggi secara keseluruhan. Dengan menerapkan teknik *Machine Learning*, terutama Metode *Decision Tree*, penelitian ini mengidentifikasi nilai rata-rata akademik (GPA) sebagai faktor utama yang mempengaruhi keberhasilan mahasiswa. Model ini mencapai tingkat akurasi sebesar 88,19% dalam mengklasifikasikan keberhasilan mahasiswa. Hasil penelitian ini menekankan pentingnya kelulusan tepat waktu serta dampak kinerja mahasiswa terhadap kualitas pendidikan tinggi secara keseluruhan. Selain itu, penelitian ini menyoroti bahwa semakin tinggi tingkat keberhasilan mahasiswa, semakin baik pula penilaian kualitas pendidikan tinggi, yang secara langsung berpengaruh pada proses akreditasi universitas. Oleh karena itu, penting untuk

mengidentifikasi mahasiswa yang berisiko putus kuliah agar dapat meningkatkan tingkat kelulusan tepat waktu, yang merupakan elemen kunci dalam evaluasi akreditasi. Ke depan, penelitian lebih lanjut dapat mengeksplorasi penggunaan teknik *Machine Learning* yang lebih canggih, seperti neural networks atau metode ensemble, guna meningkatkan model klasifikasi dalam memprediksi hasil akademik mahasiswa. Dengan demikian, universitas dapat mengembangkan alat pengambilan keputusan yang lebih andal untuk secara proaktif menangani retensi mahasiswa serta kebutuhan dukungan akademik mahasiswa.

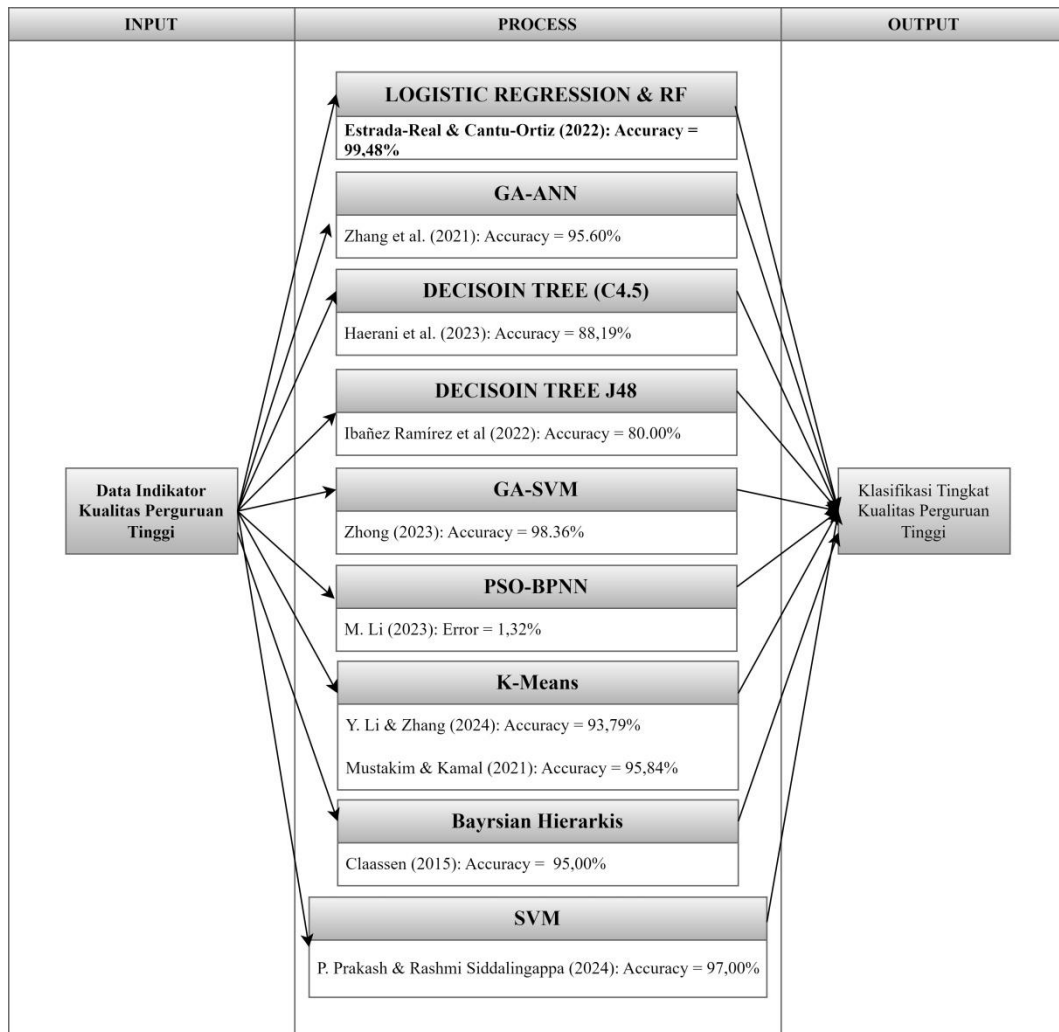
Di sisi lain, P. Prakash & Rashmi Siddalingappa, (2024) memperkenalkan model prediksi berbasis *Machine Learning* yang dirancang untuk menilai kinerja Perguruan Tinggi *Higher Education Institutions (HEIs)* di India berdasarkan kriteria yang ditetapkan oleh *National Assessment and Accreditation Council (NAAC)*. Model ini bertujuan untuk menyederhanakan proses evaluasi yang selama ini memakan waktu enam hingga tujuh bulan serta memerlukan banyak keterlibatan manusia. Dengan memanfaatkan teknologi *Machine Learning*, penilaian kualitas universitas dapat dilakukan dengan lebih efisien, mengurangi durasi dan ketergantungan pada metode tradisional yang memerlukan banyak intervensi manusia. Dalam penelitian ini, empat algoritma *Machine Learning* digunakan untuk melakukan klasifikasi multiclass terhadap prediksi skor indikator kualitatif utama, yaitu *Naive Bayes*, *Random Forest*, *K-Nearest Neighbors*, dan *Support Vector Machine*. Hasilnya menunjukkan bahwa *Support Vector Machine* memberikan akurasi tertinggi sekitar 97%, disusul oleh *Random Forest* dengan akurasi sekitar 94%. Ke depannya, penelitian lebih lanjut dapat mengeksplorasi

penggunaan algoritma *Machine Learning* lainnya di luar empat yang telah digunakan (*Naive Bayes*, *Random Forest*, *K-Nearest Neighbors*, dan *Support Vector Machine*) guna meningkatkan akurasi serta efisiensi model prediksi penilaian HEIs.

Pada studinya, Y. Li & Zhang, (2024) meneliti tentang metode pemantauan dan peningkatan kualitas pengajaran di pendidikan tinggi dengan menggunakan teknologi Big Data. Penelitian ini menggabungkan model K-Means dengan model Apriori untuk menganalisis indikator kualitas pengajaran dan menambang aturan asosiasi dari data. Tujuannya adalah untuk meningkatkan akurasi pemantauan kualitas pengajaran dan efektivitas strategi perbaikan. Model yang diusulkan menunjukkan akurasi tinggi sekitar 93,79%, serta waktu pemrosesan yang efisien sekitar 17,3 detik. Penelitian ini menyoroti potensi teknologi big data untuk memberikan wawasan bagi pembuat kebijakan pendidikan, manajer, dan pengajar, dengan harapan dapat mengoptimalkan kualitas pengajaran dan alokasi sumber daya di pendidikan tinggi.

2.2. Kerangka Teori

Pada bagian ini membahas terkait dengan pengukuran kualitas untuk menentukan parameter dan variabel sebagai data input dan untuk menyelidiki penggunaan beberapa pendekatan metode *Machine Learning* dalam pengukuran kualitas perguruan tinggi di Indonesia (PTI) yang digunakan sebagai acuan dasar dari output dalam kerangka teori. Kerangka teori untuk pengukuran perguruan tinggi di Indonesia disajikan pada Gambar 2.1.



Gambar 2. 1 Kerangka teori klasifikasi kualitas perguruan tinggi di indonesia

Parameter dalam Algoritma *Machine Learning* memiliki peran yang sangat penting dalam menciptakan model yang optimal. Parameter ini memengaruhi performa algoritma, seperti tingkat toleransi kesalahan, jumlah iterasi, dan opsi lainnya. Oleh karena itu, penting untuk mempertimbangkan kombinasi parameter yang tepat dalam menentukan algoritma *Machine Learning* yang paling sesuai.

Untuk memperoleh model terbaik yang akan di usulkan untuk klasifikasi kualitas PTI, penulis melakukan *Studi Literature* dengan mencari penelitian terkait dengan topik pembahasan tersebut. Kemudian, penulis meninjau beberapa

makalah untuk memperoleh tingkat kinerja atau performa berdasarkan nilai akurasi dan nilai *error* dari beberapa model *Machine Learning* yang digunakan. Selanjutnya, penulis mengungkap kinerja dari masing-masing model yang digunakan dengan cara mengurutkan ukuran kinerja untuk menentukan model yang digunakan sebagai landasan dasar penelitian yang akan dilakukan. Kerangka teori penelitian ini disajikan pada Tabel 2.1.

Tabel 2. 1 Kinerja Klasifikasi Kualitas menggunakan *Machine Learning*

No.	Penulis	Model	Nilai Akurasi	Nilai Error
1.	Claassen, (2015)	<i>Bayesian Hierarkis</i>	95%	-
2.	Zhang <i>et al.</i> , (2021)	<i>GA-ANN</i>	95.60%	-
3.	Mustakim & Kamal, (2021)	<i>K-Means</i>	95.84%	-
4.	Estrada-Real & Cantu-Ortiz, (2022)	<i>Logistic Regression & Random Forest</i>	99.48%	-
5.	Ibañez Ramírez <i>et al.</i> , (2022)	<i>Decision Tree J48</i>	80%	-
6.	Zhong, (2023)	<i>GA-SVM</i>	98.36%	-
7.	M. Li, (2023)	<i>PSO-BPNN</i>	-	1.32%
8.	Haerani <i>et al.</i> , (2023)	<i>Decision Tree (C4.5)</i>	88.19%	-
9.	P. Prakash & Rashmi Siddalingappa, (2024)	<i>SVM</i>	97%	-
10.	Y. Li & Zhang, (2024)	<i>K-Means</i>	93.79%	-

Tabel tersebut menunjukkan kinerja model dalam klasifikasi kualitas perguruan tinggi yang proses menggunakan berbagai model *Machine Learning* dalam mengukur kualitas universitas atau perguruan tinggi oleh para peneliti. Data yang digunakan dari masing-masing penelitian yaitu bertipe float, kategori, dan numerik sebagai variabel utama dalam mengukur kualitas pada model.

Berbagai penelitian telah dilakukan untuk mengevaluasi dan meningkatkan kualitas universitas atau perguruan tinggi dengan menggunakan model pada *Machine Learning*. Claassen (2015) menggunakan model *Bayesian Hierarkis* untuk mengukur kualitas universitas berdasarkan pemeringkatan nasional dan

internasional dengan interval kredibel sekitar 95%, menunjukkan adanya ketidakpastian dalam estimasi. Sementara itu, Mustakim & Kamal (2021) serta Haerani *et al.* (2023) menggunakan teknik klasifikasi seperti K-Means dan *Decision Tree* untuk menganalisis kualitas perguruan tinggi berdasarkan parameter akademik dan keberhasilan mahasiswa. Selain itu, penelitian oleh Estrada-Real & Cantu-Ortiz (2022) mengimplementasikan *Machine Learning* dengan algoritma yang di gunakan adalah *Logistic Regression* dan *Random Forest* dalam mengukur peringkat universitas berdasarkan data *QS World University Rankings*, mencapai akurasi tertinggi sebesar 99,48% dibandingkan dengan penggunaan model pada penelitian yang lainnya. Namun, keterbatasan penelitiannya terletak pada ketidak mampuannya mengenai perubahan dan dinamika universitas atau faktor eksternal yang dapat memengaruhi peringkat, seperti perubahan tren pendidikan global atau ketersediaan pendanaan. Hal ini dapat berdampak pada ketahanan model dalam klasifikasi kualitas perguruan tinggi atau universitas. Oleh sebab itu, Estrada-Real & Cantu-Ortiz (2022) merekomendasikan penggunaan dataset peringkat universitas lainnya pada pengembangan model sebagai langkah penelitian selanjutnya.

Di sisi lain, penelitian yang berfokus pada evaluasi kualitas pengajaran dosen menggunakan pendekatan *Machine Learning* juga menunjukkan peningkatan akurasi yang signifikan. Zhang *et al.* (2021) serta Zhong (2023) mengoptimalkan algoritma *Backpropagation* dan SVM menggunakan metode *evolutioner* seperti Algoritma Genetika (GA) untuk meningkatkan akurasi evaluasi pengajaran. Penelitian Li (2023) juga menyoroti peran *Particle Swarm*

Optimization (PSO) yang dikombinasikan dengan *Backpropagation Neural Network (BPNN)* untuk meningkatkan akurasi evaluasi kualitas pengajaran hingga 98,36%. Sementara itu, Prakash & Rashmi Siddalingappa (2024) menggunakan model SVM mencapai akurasi tertinggi sebesar 97% dalam menilai kinerja perguruan tinggi di India. Pendekatan berbasis *Machine Learning* ini menawarkan metode evaluasi yang lebih efisien dibandingkan metode tradisional, yang umumnya memakan waktu lebih lama dan bergantung pada penilaian manual.

Selain faktor akademik, penelitian juga menunjukkan bahwa aspek kehidupan mahasiswa dan faktor eksternal lainnya berpengaruh terhadap kualitas universitas. Ibañez Ramírez *et al.* (2022) menemukan bahwa lingkungan kampus, kondisi tempat tinggal, dan keadaan emosional mahasiswa menjadi faktor utama dalam menilai kualitas universitas di Kolombia Tengah, dengan model *Decision Tree* mencapai akurasi 80%. Studi oleh Y. Li & Zhang (2024) menyoroti penggunaan teknologi Big Data dalam meningkatkan pemantauan dan evaluasi kualitas pengajaran dengan algoritma K-Means dan Apriori, mencapai akurasi 93,79%. Penelitian ini menunjukkan bahwa faktor non-akademik, seperti kesejahteraan mahasiswa dan efisiensi pengelolaan data pendidikan, juga memainkan peran penting dalam meningkatkan kualitas universitas secara keseluruhan. Dengan demikian, penggabungan metode statistik, *Machine Learning*, serta analisis berbasis data besar dapat membantu universitas dalam mengambil langkah strategis guna meningkatkan reputasi akademik dan kesejahteraan mahasiswa.

Berdasarkan uraian yang sudah di bahas, menunjukkan bahwa model *Logistic Regression* dan *Random Forest* merupakan model dengan tingkat akurasi terbaik dan dapat bekerja secara efektif dalam mengukur kualitas perguruan tinggi yang memperoleh nilai akurasi sekitar 99,48% dari kedua model tersebut (Estrada-Real & Cantu-Ortiz, 2022). Selain itu. Estrada-Real & Cantu-Ortiz, (2022) merekomendasikan penggunaan dataset peringkat universitas lainnya pada pengembangan model sebagai langkah penelitian selanjutnya pada perguruan tinggi. Oleh sebab itu dalam penelitian ini penulis menyoroti kedua model tersebut, diantaranya adalah *Logistic Regression* dan *Random Forest* dalam Mengukur Kualitas Perguruan Tinggi di Indonesia.

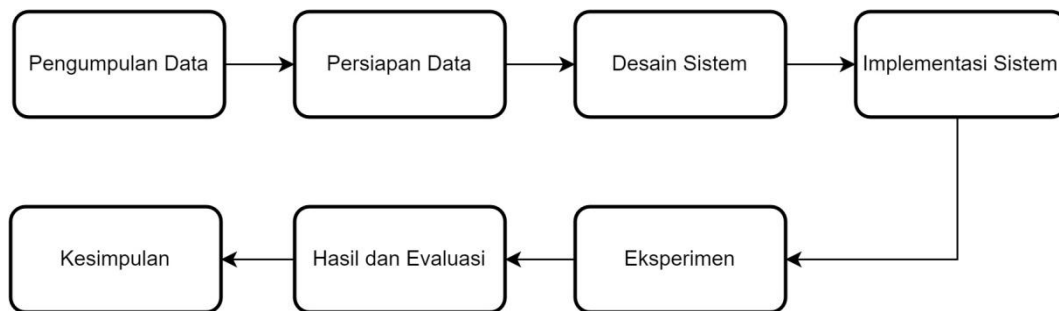
BAB III

PROSEDUR PENELITIAN

Pada bab ini dibahas prosedur penelitian serta instrumen yang digunakan untuk melakukan klasifikasi kualitas perguruan tinggi di Indonesia menggunakan metode Random Forest dan Logistic Regression. Penelitian ini menggunakan pendekatan kuantitatif karena seluruh data yang dianalisis merupakan data numerik yang diolah menggunakan teknik statistika, termasuk evaluasi model melalui confusion matrix. Secara umum, analisis isi kuantitatif merupakan teknik penelitian ilmiah yang bertujuan untuk mengidentifikasi karakteristik, pola, dan makna dari suatu konten secara sistematis. Metode ini dirancang untuk menghasilkan temuan yang objektif, valid, dan reliabel dalam proses penarikan kesimpulan (Han *et al.*, 2011)..

3.1. Prosedur Penelitian

Penulis melakukan serangkaian langkah untuk mengumpulkan data dan menjawab berbagai pertanyaan penelitian. Dalam penelitian ini, penulis membahas metode pengumpulan data, prosedur pengembangan sistem, serta eksperimen yang dilakukan. Penelitian ini bersifat eksperimental, menggunakan data penelitian yang dapat menghasilkan temuan yang terbukti melalui eksperimen atau observasi. Penelitian ini termasuk jenis penelitian eksperimental dengan menggunakan data indikator kualitas Perguruan Tinggi di Indonesia (PTI). Prosedur penelitian dijelaskan pada Gambar 3.1:



Gambar 3. 1 Prosedur penelitian

Gambar 3.1 menunjukkan bahwa tahapan penelitian ini terdiri dari delapan langkah pokok. Pertama, Studi Literatur yaitu mencari data dari buku, jurnal dan sumber-sumber lainnya yang relevan dengan topik penelitian. Kedua, Pengumpulan Data untuk proses uji coba pada sistem. Ketiga, Persiapan Data dilakukan untuk mengorganisir informasi atau data yang telah dikumpulkan sesuai dengan kebutuhan penelitian. Keempat, Desain Sistem merupakan gambaran dari proses bagaimana alur sistem dalam penelitian ini. Kelima, Implementasi Sistem merupakan inti dari program yang dilakukan secara prosedural untuk menjaga standar kualitas dan konsistensi pada sistem yang akan dirancang. Keenam, penulis melakukan Eksperimen dengan pendekatan secara ilmiah dan dataset sebagai variabel input pada penelitian terhadap model yang digunakan yaitu *Random Forest* dan *Logistic Regression*. Ketujuh, penulis melakukan Evaluasi Kinerja dengan cara menilai performa dari hasil eksperimen diantaranya berupa akurasi, f1 dan lain-lain diukur seberapa baik metode tersebut bekerja dalam mengukur kualitas perguruan tinggi berdasarkan *confusion matrix*. Tahap terakhir, selanjutnya penulis mengambil kesimpulan dari seluruh hasil penelitian.

3.1.1. Pengumpulan Data

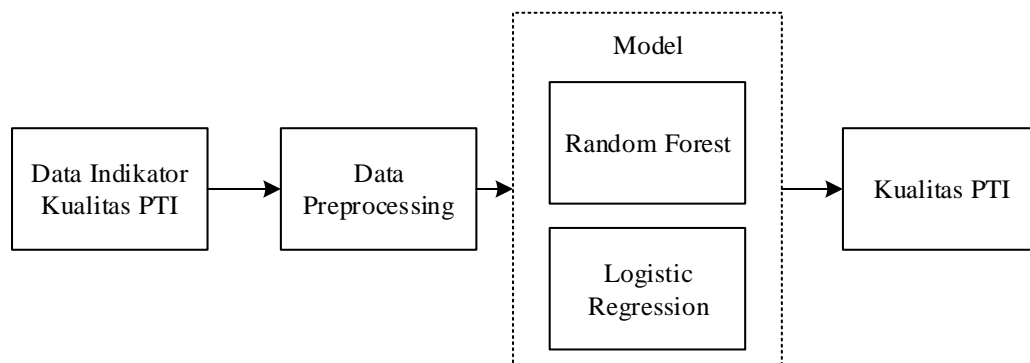
Pengumpulan data adalah proses penting yang ada di hampir setiap

penelitian. Tahap ini merupakan proses utama yang bertujuan untuk menjaga kualitas data terhadap validitas data pada penelitian tertentu, kualitas data yang buruk akan berdampak pada hasil penelitian (J. W. Creswell & Creswell, 2017). Selain itu, tahapan ini dilakukan untuk mengumpulkan setiap data yang sesuai dengan kebutuhan pada penelitian ini. Sementara itu, data yang digunakan dalam penelitian ini adalah data skunder yang bersumber dari Data Statistik Pendidikan Tinggi Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi (Dirjendikti) pada tahun 2023 yang di ambil langsung dari laman Pangkalan Data Pendidikan Tinggi/PDDIKTI (<https://pddikti.kemdiktisaintek.go.id/>). Selain itu, data yang digunakan pada penelitian ini adalah data skunder yang bersumber dari website SINTA (<https://sinta.kemdikbud.go.id/affiliations>) untuk memenuhi salah satu indikator kualitas perguruan tinggi. Selanjutnya, pengumpulan data yang dilakukan pada penelitian ini adalah mengambil dataset indikator kualitas perguruan tinggi dari sumber tersebut.

Pengambilan data dilakukan pada bulan April 2025 yang terdiri 10 atribut, yaitu Perguruan Tinggi, Skor_IKU, Program_Studi, Mahasiswa_Baru, Mahasiswa_Terdaftar, Lulusan, Dosen, Skor_Sinta_3Yr, Skor_Sinta_Overall, dan Akreditasi. Data yang diperoleh dari sumber tersebut berjumlah 118 data pada setiap atribut yang berasal dari Perguruan Tinggi Negeri Berbadan Hukum (PTN-BH), Perguruan Tinggi Negeri Badan Layanan Usaha (PTN-BLU), dan Perguruan Tinggi Negeri Satuan Kerja (PTN-Satker). Data indikator kualitas perguruan tinggi tersebut dapat dilihat pada Tabel 3.1.

3.1.3. Desain Sistem

Gambar 3.2 dalam penelitian ini menunjukkan alur proses yang terjadi di dalam sistem. Hal ini menjelaskan bagaimana sistem mengukur kualitas Perguruan Tinggi di Indonesia menggunakan *Random Forest* dan *Logistic Regression*.



Gambar 3. 2 Desain sistem

Dari gambar desain sistem 3.2 dapat di jelaskan secara singkat sebagai berikut:

- 1) Data Indikator Kualitas PTI, yaitu mengumpulkan berbagai indikator kualitas perguruan tinggi sebagai dasar analisis.
- 2) Data Preprocessing, yaitu membersihkan, menyiapkan, dan mengolah data agar siap digunakan dalam pemodelan.
- 3) Model Random Forest, yaitu membangun model klasifikasi menggunakan metode ensemble berbasis decision tree.
- 4) Model Logistic Regression, yaitu membangun model klasifikasi menggunakan pendekatan hubungan linear antara variabel.
- 5) Kualitas PTI, menghasilkan prediksi atau klasifikasi kualitas perguruan tinggi berdasarkan model yang dibangun.

A. Data Indikator Kualitas Perguruan Tinggi Indonesia

Berikut merupakan Tabel Data Indikator Kualitas Perguruan Tinggi Indonesia yang di berikan label pada masing-masing kriteria. Pada setiap kriteria terdapat rentang data yang di ambil langsung dari sumber yang telah disebutkan.

Tabel 3. 3 Data indikator kualitas perguruan tinggi Indonesia

Atribut	Rentang Data	Tipe Data	Keterangan
Perguruan_Tinggi	-	Kategorik	Nama instansi perguruan tinggi.
Skor_IKU	9,2975 Sampai 113,3688	Numerik	Skor dari indikator kinerja utama yang telah diolah oleh peneliti.
Program_Studi	4 Sampai 310	Numerik	Jumlah program studi pada perguruan tinggi.
Mahasiswa_Baru	39 Sampai 183905	Numerik	Jumlah mahasiswa terbaru pada perguruan tinggi.
Mahasiswa_Terdaftar	36 Sampai 1205711	Numerik	Jumlah mahasiswa yang menjalani perkuliahan.
Lulusan	0 Sampai 77881	Numerik	Jumlah lulusan dari perguruan tinggi.
Dosen	0 Sampai 3066	Numerik	Jumlah dosen pada perguruan tinggi.
Skor_Sinta_3Yr	593 Sampai 1230039	Numerik	Skor publikasi dalam 3 tahun terakhir pada perguruan tinggi.
Skor_Sinta_Overall	721 Sampai 4280397	Numerik	Skor publikasi secara keseluruhan waktu dari perguruan tinggi.
Akreditasi	Baik sampai unggul	Kategorik	Kategori kareditasi dari perguruan tinggi sebagai variabel target penelitian.

Pada Tabel 3.3 terdapat 10 indikator sebagai atribut data input untuk pengukuran kualitas PTI, yaitu Perguruan Tinggi, Skor_IKU, Program_Studi, Mahasiswa_Baru, Mahasiswa_Terdaftar, Lulusan, Dosen, Skor_Sinta_3Yr, Skor_Sinta_Overall, dan Akreditasi. Pada tabel tersebut terdapat kolom rentang nilai pada data yang menunjukkan nilai minimum dan maksimum pada setiap atribut dataset. Hal ini menunjukkan seberapa besar variasi data untuk setiap indikator (fitur). Ini membantu dalam Normalisasi/Standardisasi data jika

diperlukan, terutama pada algoritma sensitif seperti *Logistic Regression*. Menghindari *skewed features*, yaitu fitur dengan skala terlalu besar yang bisa mendominasi model. Menentukan *outlier*, yaitu nilai ekstrem yang bisa mempengaruhi performa model (Han *et al.*, 2011).

B. Preprocessing Data

Normalisasi *Min-Max Scaling* di gunakan untuk mengubah *value* pada data, sehingga data berada pada rentang tertentu, biasanya rentang data yang di gunakan di mulai dari angka 0 hingga 1. Teknik ini sangat berguna dalam persiapan data untuk algoritma pembelajaran mesin yang sensitif terhadap skala data (Han *et al.*, 2011).

Menurut Han *et al.*, (2011) di analogikan jika kita memiliki atribut numerik A dengan n nilai yang diamati, yaitu v_1, v_2, \dots, v_n , maka *min-max normalization* dilakukan dengan persamaan 3.1 berikut :

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} \times (\text{new max}_A - \text{new min}_A) + \text{new min}_A \quad (3.1)$$

Keterangan :

v_i = Nilai asli dari fitur A
 \min_A = Nilai minimum dari fitur A
 \max_A = Nilai maksimum dari fitur A
 new min_A = Nilai minimum baru (misalnya 0)
 new max_A = Nilai maksimum baru (misalnya 1)
 v'_i = Nilai setelah dinormalisasi

Karakteristik *Min-Max Scaling*, Pertama metode ini mempertahankan hubungan antar nilai asli, sehingga pola dalam data tidak hilang. Kedua, jika rentang data hasil dari normalisasi berada di luar rentang data asli, *Min-max*

Scalling dapat mendeteksi kesalahan atau *error*. Ketiga *Min-max Scalling* memiliki keterbatasan yaitu nilai dipengaruhi oleh *Outlier* dalam data yang dapat menyebabkan skala menjadi tidak representatif.

Berikut langkah-langkah normalisasi menggunakan *Min-Max Scalling* pada penelitian ini (Han *et al.*, 2011):

1. Identifikasi Nilai Minimum dan Maksimum
2. Tentukan Rentang Baru
3. Terapkan Rumus Min-Max Scalling
4. Terapkan pada semua nilai
5. Verifikasi dan simpan hasilnya

3.1.4. Proses Utama

Langkah-Langkah Analisis Menggunakan Model *Random Forest* dan *Logistic Regression* untuk Klasifikasi:

1. Tentukan data input (Dataset)

Siapkan data yang akan dianalisis, lengkap dengan fitur (variabel input) dan label kelas (kategori hasil) sebagai target klasifikasi.

2. Pilih fitur (atribut) yang relevan (*Random Forest*)

Identifikasi atribut yang paling mempengaruhi kelas target. Bisa dibantu dengan teknik seleksi fitur.

3. Identifikasi fitur yang berpengaruh (*Logistic Regression*)

Tentukan fitur yang akan digunakan untuk memisahkan data ke dalam dua kelas.

4. Bagi data menjadi data latih dan data uji

Misalnya 80% untuk melatih model dan 20% untuk menguji kemampuan klasifikasinya.

5. Latih model *Random Forest*

Model akan membangun banyak pohon keputusan dari data latih untuk menentukan kategori dari setiap data.

6. Latih model *Logistic Regression*

Model akan belajar membentuk garis pemisah antar kelas dengan menghitung bobot setiap fitur.

7. Lakukan klasifikasi pada data uji (*Random Forest*)

Model mengelompokkan data ke dalam kelas berdasarkan hasil voting dari seluruh pohon.

8. Lakukan klasifikasi pada data uji (*Logistic Regression*)

Model mengklasifikasikan data ke dalam kelas berdasarkan nilai probabilitas.

9. Evaluasi performa klasifikasi

Gunakan metrik klasifikasi seperti *Confusion Matrix*.

10. Tarik kesimpulan dari hasil klasifikasi

Analisis akurasi klasifikasi, fitur penting yang digunakan model, dan cocokkan dengan kebutuhan aplikasi atau kasus nyata.

3.1.5. Implementasi Sistem

Setiap model *Random Forest* terdiri dari kumpulan pohon keputusan (*Decision Trees*) yang dibangun secara acak dari subset data pelatihan. Setiap

pohon dalam *Random Forest* memberikan prediksi, dan hasil akhir ditentukan berdasarkan mayoritas suara (*voting*) dari seluruh pohon (untuk klasifikasi) atau rata-rata (untuk regresi). Metode ini mengurangi risiko *overfitting* yang umum terjadi pada pohon keputusan tunggal dengan cara menggabungkan hasil dari banyak pohon yang berbeda. *Random Forest* efektif dalam menangani data yang kompleks dan bekerja dengan baik meskipun terdapat korelasi antar fitur.

Sementara itu, *Logistic Regression* digunakan untuk memprediksi probabilitas dari sebuah kelas atau kategori tertentu. Metode ini sangat cocok untuk masalah klasifikasi biner (dua kelas), seperti menentukan apakah email adalah spam atau bukan. *Logistic Regression* menggunakan fungsi logistik (*sigmoid*) untuk mengubah keluaran linier menjadi nilai probabilitas antara 0 dan 1. Jika nilai probabilitas tersebut melewati ambang batas tertentu (misalnya 0.5), maka model mengklasifikasikan input ke dalam salah satu kategori.

3.1.6. Eksperimen

Tahap uji coba dalam penelitian ini dilakukan dengan menggunakan program alat *Kaggle Editor* dengan Bahasa pemrograman *Python* untuk mengukur kualitas perguruan tinggi di Indonesia menggunakan Model *Random Forest* dan *Logistic Regression*. Setelah diperoleh hasil klasifikasi dari model tersebut, kemudian dievaluasi menggunakan tabel *Confusion Matrix*. Selanjutnya, akan diambil kesimpulan berdasarkan hasil analisis yang diperoleh. Berikut pada tabel 3.4 merupakan komposisi data latih dan data uji yang akan di lakukan:

Tabel 3. 4 Komposisi data latih dan data uji

Metode1	Metode2	Model	Rasio Data	Data Latih	Data Uji	Nama Pelatihan Metode1	Nama Pelatihan Metode2
Random Forest	Logistik Regression	A	90:10	106	12	RF-A	LR-A
		B	80:20	94	24	RF-B	LR-B
		C	70:30	82	36	RF-C	LR-C
		D	60:40	70	48	RF-D	LR-D
		E	50:50	58	60	RF-E	LR-E

Dari Tabel tersebut menunjukkan komposisi dari 118 data secara keseluruhan dari masing-masing atribut yang dibagi menjadi lima kelompok dengan komposisi data latih dan data uji yang model A sebesar 90% : 10%, model B sebanyak 80% : 20%, model C sebesar 70% : 30%, model D sebanyak 60% : 40%, dan model E sebesar 50% : 50%.

Setelah proses pelatihan dan pengujian dilakukan pada setiap model, tahap berikutnya adalah mengevaluasi kinerjanya. Evaluasi dilakukan menggunakan *Confusion Matrix* yang merupakan teknik evaluasi yang luas digunakan dalam bidang machine learning untuk mengukur akurasi klasifikasi. Teknik evaluasi ini digunakan untuk mengukur nilai *Accuracy*, *Precision*, *Recall*, *F-Measure*, *True Positive Rate (TPR)*, dan *False Positive Rate (FPR)*. Berikut merupakan konsep dari *Confusion Matrix* yang ditunjukkan pada Tabel 3.5.

Tabel 3. 5 Konsep *confusion matrix*

Confusion Matrix		Kelas Aktual	
		Positif	Negatif
Kelas Prediksi	Positif	TP	FP
	Negatif	FN	TN

Selanjutnya, metode *Confusion Matrix* ini digunakan untuk perbandingan antara data hasil klasifikasi dengan data aktual (Wiraswendro & Soetanto, 2022). Pengujian dilanjutkan dengan menghitung nilai *Accuracy*, *Precision*, *Recall*, *F-Measure*, *TPR* dan *FPR*. *Accuracy* digunakan untuk mengetahui seberapa akurat

model dapat mengklasifikasikan data target dengan benar. *Precision* menggambarkan sejauh mana keakuratan data yang diminta dengan hasil prediksi model. *Recall* menunjukkan tingkat keberhasilan sistem dalam mengidentifikasi semua kasus dan menemukan kembali suatu informasi. F-measure merupakan *metrix* gabungan dari *precision* dan *recall*. *True Positive Rate (TPR)* bisa disebut sebagai *Recall* atau *Sensitivity* mengukur sejauh mana model mengidentifikasi kasus positif yang diklasifikasikan dengan benar sebagai kasus positif. Sedangkan *False Positive Rate (FPR)* mengukur seberapa sering model salah mengidentifikasi kelas negatif yang diklasifikasikan positif (Han *et al.*, 2011).

$$\text{Accuracy} = \frac{\text{Jumlah klasifikasi benar}}{\text{Jumlah dokumen uji}} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\%$$

$$\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{TPR} = \frac{TP}{TP+FN} \times 100\%$$

$$\text{FPR} = \frac{FP}{FP+TN} \times 100\% \quad (3.8)$$

Keterangan:

- True Positive (TP): Kelas yang sebenarnya positif, dan diprediksi menjadi positif. Misalnya, akreditasi sebenarnya “baik”, dan diprediksi sebagai akreditasi “baik”.
- False Positive (FP): Kelas yang sebenarnya negatif, dan diprediksi menjadi positif. Misalnya, akreditasi sebenarnya “unggul”, dan diprediksi menjadi akreditasi “baik”.
- False Negative (FN): Kelas yang sebenarnya positif, namun diprediksi menjadi negatif. Misalnya, akreditasi sebenarnya “baik”, namun diprediksi sebagai akreditasi “unggul”.
- True Negative (TN): Kelas yang sebenarnya negatif, dan diprediksi menjadi negatif. Misalnya, akreditasi sebenarnya bukan “baik”, dan diprediksi sebagai bukan akreditasi “baik”.

Selain metrik utama, penelitian ini juga menggunakan Area Under the Curve (AUC) dari Receiver Operating Characteristic (ROC) untuk menilai kemampuan model dalam membedakan kelas positif dan negatif. Nilai AUC berkisar antara 0 hingga 1, di mana nilai yang lebih mendekati 1 menunjukkan performa klasifikasi yang lebih baik. AUC menjadi metrik pelengkap yang penting karena memberikan gambaran menyeluruh tentang keseimbangan antara sensitivitas dan spesifisitas, terutama pada dataset dengan distribusi kelas tidak seimbang.

$$AUC = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \times \frac{TPR_{i+1} + TPR_i}{2} \quad (3.9)$$

Keterangan:

- *AUC* (Area Under the Curve): Luas di bawah kurva ROC (Receiver Operating Characteristic), digunakan untuk mengukur kemampuan model dalam membedakan antara kelas positif dan negatif. Nilai AUC berkisar antara 0 hingga 1, dengan nilai mendekati 1 menunjukkan performa klasifikasi yang sangat baik.
- (FPR_i) (False Positive Rate ke-i): Rasio antara jumlah prediksi positif yang salah (false positive) terhadap total data negatif aktual. Dihitung sebagai:

$$FPR = \frac{FP}{FP + TN} \quad (3.10)$$

- (TPR_i) (True Positive Rate ke-i): Juga dikenal sebagai sensitivitas atau recall, yaitu rasio antara jumlah prediksi positif yang benar (true positive) terhadap total data positif aktual. Dihitung sebagai:

$$TPR = \frac{TP}{TP + FN} \quad (3.11)$$

- n : Jumlah titik (threshold) yang digunakan untuk membentuk kurva ROC.

ROC curve adalah grafik yang menunjukkan trade-off antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai nilai ambang (*threshold*) dari probabilitas klasifikasi dengan persamaan dibawah ini: Nilai TPR menunjukkan seberapa banyak data positif yang berhasil diklasifikasikan dengan benar oleh model. Sedangkan nilai FPR menunjukkan seberapa banyak data negatif yang secara salah diklasifikasikan sebagai positif. Dengan memvariasikan

ambang (*threshold*) dari 0 sampai 1, kita bisa mendapatkan banyak pasangan nilai (TPR & FPR), yang kemudian digambarkan menjadi kurva ROC (Dalton, 2015).

Menurut Muschelli III, (2020) AUC adalah luas di bawah kurva ROC, yang mengukur kemampuan model untuk membedakan kelas positif dan negative dengan interpretasi nilai yang disajikan pada Tabel 3.5.

Tabel 3. 6 Interpretasi nilai AUC

Nilai AUC	Interpretasi
0.90 – 1.00	Sangat baik (Excellent)
0.80 – 0.90	Baik (Good)
0.70 – 0.80	Cukup (Fair)
0.60 – 0.70	Buruk (Poor)
0.50 – 0.60	Sangat buruk / tidak lebih baik dari random

Dengan menggabungkan hasil dari Confusion Matrix dan skor AUC, penilaian terhadap kinerja model dalam penelitian ini menjadi lebih menyeluruh dan akurat, serta dapat digunakan sebagai dasar dalam memilih model terbaik untuk prediksi yang dilakukan.

3.2. Instrumen Penelitian

Pada Tabel 3.7 menampilkan daftar parameter yang digunakan dalam penelitian ini, mencakup variabel independen atau terikat, variabel intervening atau penghubung, serta variabel dependen atau bebas. Fokus utama penelitian ini adalah pada variabel independen, yang juga disebut sebagai variabel terikat. Selain itu, variabel terikat merupakan variabel yang dipengaruhi atau merupakan hasil dari variabel bebas. Dalam penelitian ini, nilai *error* berfungsi sebagai variabel terikat.

Tabel 3. 7 Variabel penelitian

Varibel Independen	Main Process	Variabel Intervening	Variabel Dependen
Data Indikator Kualitas Perguruan Tinggi	<i>Random Forest</i> dan <i>Logistic Regression</i>	Pengukuran Kualitas PTI	Nilai Akurasi

Fokus utama penelitian adalah variabel independen, juga dikenal sebagai variabel terikat. Selain itu, variabel terikat adalah variabel yang dipengaruhi atau akibat dari variabel bebas. Dalam penelitian ini, nilai *error* adalah variabel terikat.

Secara teori, variabel intervening, atau variabel penghubung, mempengaruhi hubungan antara variabel bebas dan variabel terikat secara tidak langsung, tidak bisa diamati atau diukur secara langsung. Variabel intervening berada di antara variabel independen dan dependen, sehingga tidak secara langsung memengaruhi perubahan atau kemunculan variabel dependen.

BAB IV

KLASIFIKASI KUALITAS PERGURUAN TINGGI DENGAN *RANDOM FOREST*

Random Forest pertama kali diperkenalkan oleh Breiman pada tahun 2001. Menurut Pamuji & Ramadhan (2021) *Random Forest* memiliki beberapa kelebihan yakni memiliki performa baik dalam melakukan klasifikasi, menghasilkan tingkat eror yang relatif rendah, bekerja secara efisien dalam data dengan jumlah besar, dan efektif dalam menangani *missing data*.

Metode *Random Forest* adalah salah satu jenis metode dari *ensemble learning* yang biasa digunakan untuk pengklasifikasian dan regresi (Ramadhan *et al.*, 2022). *Ensemble Learning* yakni menggabungkan beberapa pengklasifikasian untuk meningkatkan kinerja model. Menurut Dietterich dalam Willy *et al.* (2021) pada *ensemble learning*, pembelajaran dilakukan dengan voting dari sekumpulan hipotesa yang telah dibuat. *Output class* dipilih dari hipotesa dengan vote tertinggi.

Menurut Sadewo dalam Primajaya & Sari (2018) terdapat tiga aspek penting dalam metode *Random Forest* yakni:

- 1) Membangun pohon keputusan dengan teknik *Bootstrap Sampling*
- 2) Setiap pohon keputusan melakukan prediksi secara acak
- 3) *Random Forest* melakukan prediksi dengan proses *voting*. Untuk klasifikasi menggunakan nilai mayoritas atau nilai yang paling sering muncul, sedangkan untuk regresi dilakukan dengan menggunakan nilai rata-rata.

Model dalam *Random Forest* dibangun dengan menggunakan algoritma pohon keputusan atau (*Decision Tree*). Algoritma yang dapat digunakan yakni ID3 (*Iterative Dichotomiser 3*) yang mana dalam pemisahan simpul didasarkan pada nilai *entropy*, yang mencerminkan tingkat ketidakpastian atau impuritas dari atribut. *Entropy* yang tinggi menunjukkan bahwa data sangat tercampur, sedangkan *entropy* rendah berarti data lebih homogen. Kedua adalah Algoritma CART (*Classification and Regression Tree*) berdasarkan nilai *Gini Index* untuk membagi simpul (Ashfania *et al.*, 2023). Febriansyah *et al.* (2023) mendefinisikan *Gini Index* mengukur tingkat ketidakseimbangan distribusi kelas dalam satu kelompok data. Semakin kecil nilai *Gini*, semakin baik pemisahan kelas yang terjadi.

Berikut rumus pemisahan simpul didasarkan pada pencarian *entropy* yang ditunjukkan pada rumus 4.1.

$$Entropy(S) = \sum_{i=1}^n -P_i * \log_2 P_i \quad (4.1)$$

Di mana nilai S merupakan himpunan kasus, n adalah jumlah partisi S , dan P_i adalah proporsi himpunan kasus ke- i terhadap himpunan kasus.

Selanjutnya dilakukan pencarian *Information Gain*. *Information Gain* merupakan perolehan informasi menggunakan *Entropy* untuk menentukan atribut terbaik (Norhalimi & Siswa, 2022). Berikut rumus persamaan pencarian *information gain* yang ditujukan pada rumus 4.2.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (4.2)$$

Di mana S merupakan himpunan kasus, A sebagai atribut, n merupakan jumlah partisi atribut A , $|S_i|$ adalah proporsi S_i terhadap S , $|S|$ adalah jumlah kasus dalam S , dan $entropy(S_i)$ merupakan $entropy$ untuk sampel memiliki nilai ke- i (Suntoro, 2019).

Setelah *Information Gain* sudah di hitung untuk setiap atribut, langkah selanjutnya adalah memilih atribut dengan nilai *Gain* terbesar untuk dijadikan atribut pemisah pada pohon keputusan. Secara konsep, rumus sederhananya pada persamaan 4.3 (Suntoro, 2019) :

$$\text{Atribut terbaik} = \arg \max_{A \in \text{atribut}} \text{Gain}(A) \quad (4.3)$$

Keterangan :

- **arg max** berarti “nilai argumen yang menghasilkan nilai maksimum”.
- **Gain(A)** dari semua atribut A , kita pilih atribut yang memiliki nilai $Gain(A)$ paling besar untuk di jadikan node pada pohon keputusan.
- **$A \in \text{atribut}$** artinya pemilihan dilakukan dari seluruh atribut yang ada pada dataset.

Berikut algoritma *Random Forest* menurut (Jackins *et al.*, 2021).

1. Pilih fitur secara acak sebanyak “ n ” dari total “ k ” fitur. $n < k$
2. Hitung simpul fitur “ n ” menggunakan titik pemisah terbaik
3. Kategorikan simpul kedalam simpul anak dengan pemisahan terbaik
4. Ulangi langkah 1 sampai 3 sampai jumlah simpul tercapai
5. Ulangi langkah 1 sampai 4 sebanyak “ n ” untuk membuat “ n ” pohon

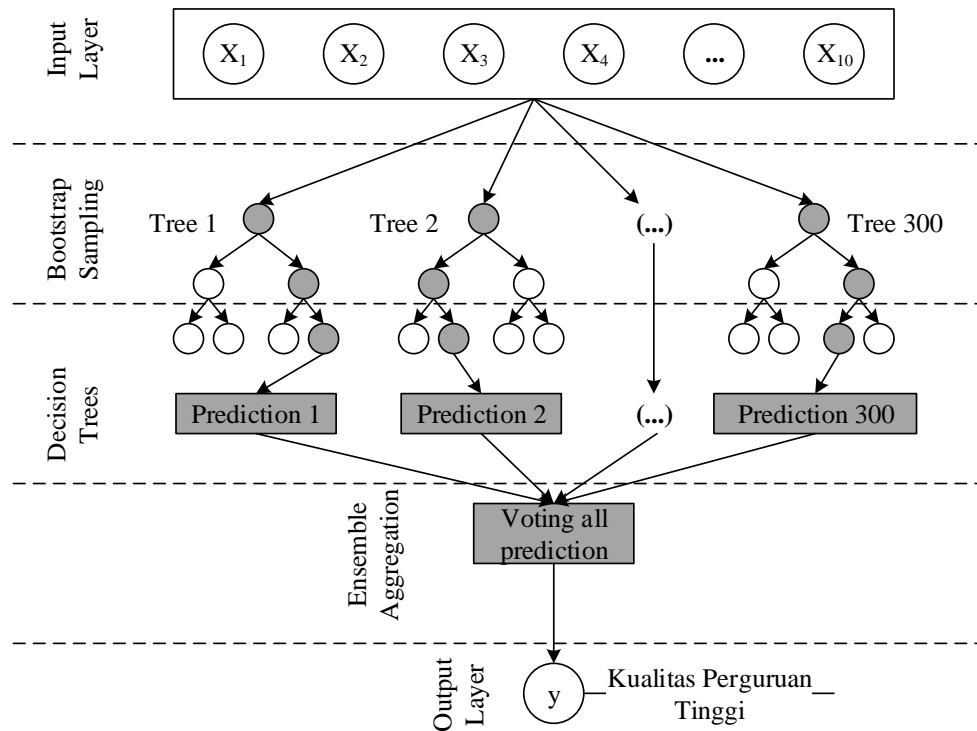
Pemodelan sistem menggunakan algoritma *Random Forest* telah berhasil diterapkan dalam penelitian ini untuk mengukur kualitas perguruan tinggi di Indonesia. Model ini dipilih karena kemampuannya yang unggul dalam mengelola

data berskala besar dengan banyak variabel serta efektivitasnya dalam menangani klasifikasi yang kompleks (Nugraha & Syarif, 2023). Hasil dari pemodelan menunjukkan performa yang baik dalam mengklasifikasikan kualitas perguruan tinggi. Penjelasan lebih detail mengenai kinerja model serta interpretasi hasilnya akan dibahas dan ditunjukkan pada sub-bab selanjutnya.

4.1. Desain Model *Random Forest*

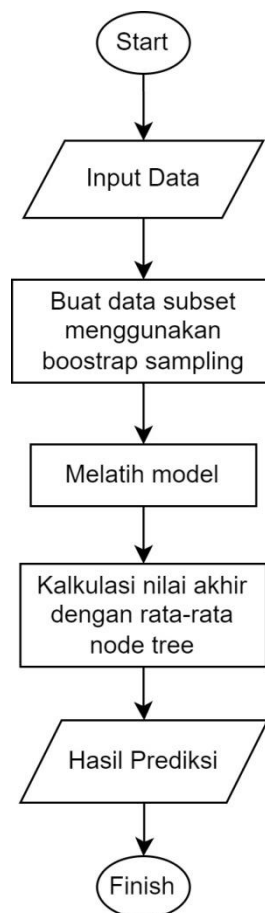
Desain penelitian ini menjelaskan tahapan-tahapan yang ditempuh peneliti untuk memperoleh hasil yang ditujukan sebagai jawaban atas permasalahan yang diangkat. Pada penelitian ini digunakan model *Random Forest* (RF) dengan skema pelatihan mulai dari RF-A hingga RF-E untuk melakukan klasifikasi tingkat akreditasi dalam rangka mengukur kualitas perguruan tinggi di Indonesia. Model RF yang dibangun memanfaatkan 300 pohon keputusan (Tree), sehingga diharapkan mampu memberikan prediksi yang lebih konsisten, stabil, dan akurat. Gambaran lebih rinci mengenai rancangan arsitektur model *Random Forest* ini dapat dilihat pada Gambar 4.1.

Proses training dari model *Random Forest* (RF) dimulai dengan memanggil data training yang berisi indikator-indikator kualitas perguruan tinggi yang telah disiapkan sebelumnya. Pada tahap awal, ditentukan parameter utama yaitu jumlah pohon keputusan (tree) sebanyak 300 atau nilai $n_estimator = 300$. Setiap pohon akan dibangun dengan mengambil subset data dan subset fitur secara acak untuk memastikan keragaman dalam proses pembentukan model.



Gambar 4. 1 Arsitektur model *Random Forest* 300 tree

Selanjutnya, setiap pohon keputusan melakukan proses pembelajaran secara independen terhadap subset data yang diperolehnya, menghasilkan prediksi parsial untuk klasifikasi tingkat akreditasi perguruan tinggi. Hasil prediksi dari seluruh pohon kemudian digabungkan menggunakan mekanisme voting mayoritas guna menghasilkan keputusan akhir yang lebih akurat dan stabil. Proses training akan terus berlangsung hingga jumlah pohon yang terbentuk mencapai 300 atau model dianggap mencapai titik konvergensi dalam hal kestabilan performa. Setelah seluruh pohon terbentuk, model *Random Forest* yang telah terlatih akan menyimpan struktur pohon-pohon keputusan tersebut untuk digunakan pada tahap pengujian (*testing*). *Flowchart* klasifikasi *Random Forest* kualitas perguruan tinggi ini ditunjukkan pada Gambar 4.2.



Gambar 4. 2 Flowchart Model *Random Forest*

4.2. Implementasi Model *Random Forest*

Dalam penelitian ini, algoritma *Random Forest* digunakan untuk membangun model klasifikasi kualitas perguruan tinggi di Indonesia. Pemilihan *Random Forest* bukan tanpa alasan algoritma ini dikenal andal dalam mengelola data dengan dimensi dan variabel yang kompleks, serta mampu memberikan hasil klasifikasi yang akurat dan tahan terhadap overfitting. Pada bagian ini dijelaskan proses implementasi model klasifikasi, dimulai dari pengolahan data menggunakan Kaggle Editor dan pemrograman Python, hingga menghasilkan output berupa klasifikasi kualitas perguruan tinggi. Model yang dipilih, yaitu *Random Forest*, merupakan algoritma berbasis ensemble learning dari pohon

keputusan (decision tree) yang terbukti efektif untuk jenis data tabular seperti data pemeringkatan, menjadikannya pilihan tepat dalam konteks penelitian ini.

Sebelum menjalankan model, data terlebih dahulu dibagi menjadi dua bagian, yaitu data latih dan data uji. Langkah ini penting dilakukan untuk mencegah terjadinya overfitting serta memastikan bahwa model yang dibangun mampu melakukan generalisasi dengan baik terhadap data perguruan tinggi yang belum pernah dilihat sebelumnya. Setelah itu, model *Random Forest* diinisialisasi dengan beberapa parameter kunci, seperti jumlah pohon dalam hutan (*n_estimators*), kedalaman maksimum setiap pohon (*max_depth*), dan jumlah minimum sampel untuk membagi simpul (*min_samples_split*).

Berbeda dengan pendekatan regresi yang memprediksi nilai kontinu, *Random Forest* untuk klasifikasi bekerja dengan membangun sekumpulan pohon keputusan dan menentukan kelas mayoritas dari hasil prediksi setiap pohon. Proses ini secara efektif mengurangi varians dan meningkatkan akurasi dibandingkan dengan satu pohon keputusan tunggal. Untuk menangani potensi ketidakseimbangan data antar kelas misalnya, lebih banyak perguruan tinggi di klaster "Baik" daripada "Unggul", digunakan parameter *class_weight = 'balanced'*. Rincian lengkap mengenai parameter dan konfigurasi model *Random Forest* yang digunakan dalam pelatihan ditampilkan pada Tabel 4.1.

Model *Random Forest* pada penelitian ini dirancang untuk menyelesaikan masalah klasifikasi dengan menentukan tiga kategori kualitas perguruan tinggi di Indonesia yaitu, "Unggul", "Baik Sekali", dan "Baik" berdasarkan berbagai

indikator. Jumlah pohon keputusan yang digunakan dalam model ini sebanyak 300 ($n_estimators=300$). Setiap pohon memberikan "suara" untuk klasifikasi, dan hasil akhir ditentukan oleh suara mayoritas, sehingga model menjadi semakin kuat dan akurat.

Tabel 4. 1 Parameter model *Random Forest*

Parameter	Nilai	Keterangan	Penjelasan
<code>n_estimators</code>	300	Jumlah Pohon	Menentukan jumlah pohon keputusan (decision tree) yang akan dibangun dalam model.
<code>max_depth</code>	10	Kedalaman Maksimum Pohon	Membatasi kedalaman maksimum dari setiap pohon.
<code>min_samples_split</code>	5	Jumlah Minimum Sampel untuk Membagi Simpul	Menentukan jumlah minimum sampel data yang harus ada di sebuah simpul (node) agar simpul tersebut dapat dibagi menjadi dua.
<code>min_samples_leaf</code>	1	Jumlah Minimum Sampel pada Daun	Menentukan jumlah minimum sampel data yang harus ada di setiap daun (leaf node) setelah pembagian.
<code>class_weight</code>	'balanced'	Bobot Kelas	Mengatur bobot untuk setiap kelas secara otomatis berdasarkan jumlah sampelnya.
<code>random_state</code>	42	Pengaturan Angka Acak	Memastikan bahwa hasil yang didapat akan selalu sama setiap kali model dijalankan dengan data yang sama.

Kedalaman maksimum setiap pohon dibatasi hingga 10 ($max_depth=10$) untuk mencegah model menjadi terlalu kompleks dan kehilangan kemampuan generalisasi. Parameter $min_samples_split=5$ dan $min_samples_leaf=1$ mengontrol kondisi percabangan pohon untuk memastikan model tidak terlalu sensitif terhadap data latih. Selain itu, model ini menggunakan angka acak tetap ($random_state=42$) supaya hasil yang diperoleh konsisten dan dapat direproduksi jika dilakukan pelatihan ulang. Dengan pengaturan ini, diharapkan model *Random Forest* mampu memberikan klasifikasi kualitas perguruan tinggi yang andal dan membantu para pemangku kepentingan dalam pengambilan keputusan yang lebih objektif.

Dalam penelitian ini, proses pelatihan model *Random Forest* dilakukan dengan membagi data awal menjadi dua bagian utama, yaitu data pelatihan dan data pengujian. Pembagian ini bertujuan untuk memastikan bahwa model dapat belajar secara maksimal dari sebagian data, lalu diuji keakuratannya menggunakan data lain yang belum pernah dilihat sebelumnya. Hal ini penting agar performa model mencerminkan kemampuannya dalam menghadapi data baru di dunia nyata.

Tabel 4. 2 Distribusi data pelatihan model *Random Forest*

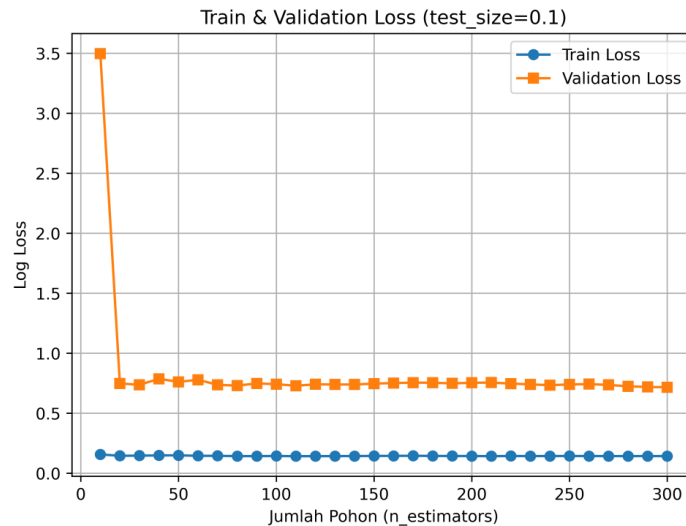
Nama Pelatihan	Rasio Data	Data Latih	Data Uji
RF-A	90:10	106	12
RF-B	80:20	94	24
RF-C	70:30	82	36
RF-D	60:40	70	48
RF-E	50:50	58	60

Untuk mengeksplorasi bagaimana pembagian data memengaruhi hasil pelatihan pada Tabel 4.2, digunakan beberapa variasi proporsi antara data pelatihan dan pengujian. Setiap variasi tersebut diberi kode khusus, yaitu RF-A, RF-B, RF-C, RF-D, dan RF-E, yang masing-masing menunjukkan perbedaan dalam proporsi pembagian data. Model-model ini kemudian dilatih untuk mengukur kualitas perguruan tinggi berdasarkan tingkat Akreditasi, yang menjadi variabel target penelitian, dengan tiga kategori Akreditasi yaitu Baik, Baik Sekali, dan Unggul.

4.2.1. Pelatihan Model RF-A

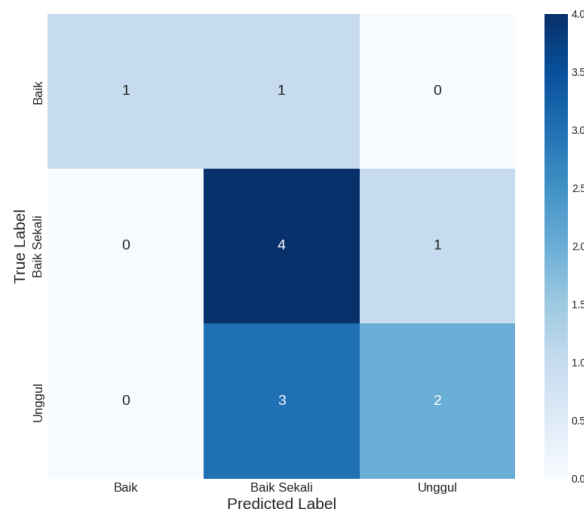
Pada proses eksperimen dari pelatihan model RF-A dalam klasifikasi kualitas perguruan tinggi di Indonesia digunakan proporsi data latih sebesar 90% atau sekitar 106 data dan data uji sebesar 10% atau sekitar 12 data. Model dilatih

dengan jumlah $n_estimators$ yang telah ditentukan, yaitu sebanyak 300 pohon. Hasil pelatihan menunjukkan nilai *train loss* sekitar 0.1422 dan nilai *validation loss* sekitar 0.7164 sebagaimana ditampilkan pada Gambar 4.3.



Gambar 4. 3 Grafik train dan validation loss model RF-A

Pada Gambar 4.3 menunjukkan nilai *train* dan *validation loss* pada proses pelatihan model RF-A yang terus mengalami penurunan hingga mencapai titik konvergensi pada $n_estimator$ ke-300, sehingga model tersebut dianggap telah optimal. Selanjutnya, model optimal yang telah didapatkan, kemudian digunakan untuk mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi dengan tiga (3) klasifikasi yaitu “Baik”, “Baik Sekali”, dan “Unggul”. Hasil dari pengujian model RF-A dalam melakukan klasifikasi tingkat akreditasi perguruan tinggi disajikan pada Gambar 4.4.



Gambar 4. 4 Hasil klasifikasi model RF-A terhadap tingkat akreditasi

Hasil klasifikasi model RF-A terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 4.4 menunjukkan performa klasifikasi model RF-A terhadap 12 data uji pada tiga kategori akreditasi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model hanya benar pada 1 dari 2 data, menandakan kesulitan membedakan dari kelas tengah. Pada kategori Baik Sekali, model relatif baik dengan 4 dari 5 data tepat, meskipun 1 salah diprediksi sebagai Unggul. Sementara pada kategori Unggul, model kurang optimal karena hanya 2 dari 5 data yang benar, dengan mayoritas salah diklasifikasikan sebagai Baik Sekali. Hal ini menunjukkan kecenderungan model mengalami bias ke kelas tengah.

Secara eksplisit, model cenderung memusatkan prediksi pada kategori Baik Sekali sebagai posisi tengah ketika menghadapi ketidakpastian, sehingga akurasi pada kelas Baik dan Unggul menurun. Pola bias ini menjadi tantangan dalam klasifikasi kategori yang bersifat ordinal seperti kualitas perguruan tinggi. Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-*

Score maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 4.3.

- True Positive (TP): Kelas yang sebenarnya positif, dan diprediksi menjadi positif. Misalnya, akreditasi sebenarnya “baik”, dan diprediksi sebagai akreditasi “baik”.
- False Positive (FP): Kelas yang sebenarnya negatif, dan diprediksi menjadi positif. Misalnya, akreditasi sebenarnya “unggul”, dan diprediksi menjadi akreditasi “baik”.
- False Negative (FN): Kelas yang sebenarnya positif, namun diprediksi menjadi negatif. Misalnya, akreditasi sebenarnya “baik”, namun diprediksi sebagai akreditasi “unggul”.
- True Negative (TN): Kelas yang sebenarnya negatif, dan diprediksi menjadi negatif. Misalnya, akreditasi sebenarnya bukan “baik”, dan diprediksi sebagai bukan akreditasi “baik”.

Nama Pelatihan	Rasio Data	Data Latih	Data Uji
RF-A	90:10	106	12
RF-B	80:20	94	24
RF-C	70:30	82	36
RF-D	60:40	70	48
RF-E	50:50	58	60

Tabel 4. 3 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-A

Kategori	TP	TN	FP	FN
Baik	1	10	0	1
Baik Sekali	4	4	3	1
Unggul	3	6	1	2

Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang digunakan untuk mengetahui performa model RF-A dalam mengukur kualitas

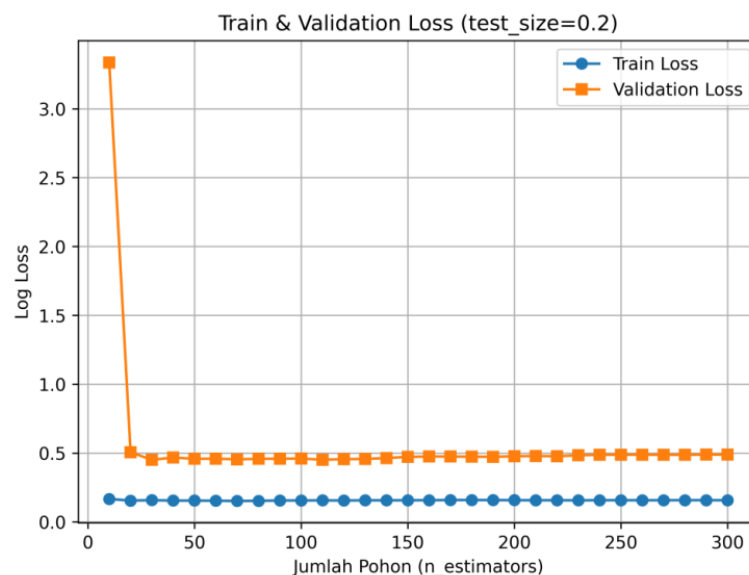
perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 4.4.

Tabel 4. 4 Performa hasil pengujian model RF-A

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Baik	0.58	1	0.5	0.67	2
Baik Sekali		0.5	0.8	0.62	5
Unggul		0.67	0.4	0.5	5

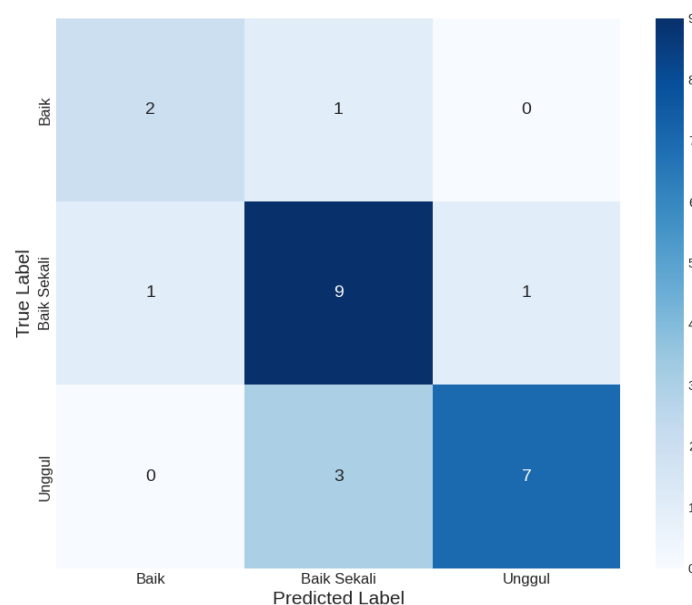
4.2.2. Pelatihan Model RF-B

Pada proses eksperimen dari pelatihan model RF-B dalam klasifikasi kualitas perguruan tinggi di Indonesia digunakan proporsi data latih sebesar 80% atau sekitar 94 data dan data uji sebesar 20% atau sekitar 24 data. Model dilatih dengan jumlah pohon keputusan ($n_estimators$) yang telah ditentukan, yaitu sebanyak 300 pohon. Hasil pelatihan menunjukkan nilai *train loss* sekitar 0.1570 dan nilai *validation loss* sekitar 0.4882 sebagaimana ditampilkan pada Gambar 4.5 berikut ini:



Gambar 4. 5 Grafik train dan validation loss model RF-B

Pada Gambar 4.5 menunjukkan nilai *train* dan *validation loss* pada proses pelatihan model RF-B yang terus mengalami penurunan hingga mencapai titik konvergensi pada *n_estimator* ke-300, sehingga model tersebut dianggap telah optimal. Selanjutnya, model optimal yang telah didapatkan, kemudian digunakan untuk mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi dengan tiga (3) klasifikasi yaitu “Baik”, “Baik Sekali”, dan “Unggul”. Hasil dari pengujian model RF-B dalam melakukan klasifikasi tingkat akreditasi perguruan tinggi disajikan pada Gambar 4.6.



Gambar 4. 6 Hasil klasifikasi model RF-B terhadap tingkat akreditasi

Hasil klasifikasi model RF-B terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 4.6 memperlihatkan hasil klasifikasi model RF-B terhadap 24 data uji dalam tiga kategori kualitas perguruan tinggi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model mampu mengklasifikasikan 2 dari 3 data dengan benar, meskipun masih terdapat 1 data yang salah diprediksi sebagai Baik Sekali. Untuk

kategori Baik Sekali, performa model terlihat lebih stabil dengan 9 dari 11 data berhasil diklasifikasikan secara tepat, sementara 2 data lainnya salah diprediksi ke kategori Baik maupun Unggul. Adapun pada kategori Unggul, model berhasil mengenali 7 dari 10 data, namun 3 data lainnya salah diklasifikasikan ke kategori Baik Sekali.

Singkatnya, hasil model RF-B menunjukkan bahwa meskipun model cukup andal dalam mengenali setiap kelas, masih terdapat tantangan dalam membedakan karakteristik yang berdekatan antar kategori, terutama antara Baik Sekali dan Unggul. Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-Score* maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 4.5.

Tabel 4. 5 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-B

Kategori	TP	TN	FP	FN
Baik	2	20	1	1
Baik Sekali	9	10	3	2
Unggul	8	13	1	2

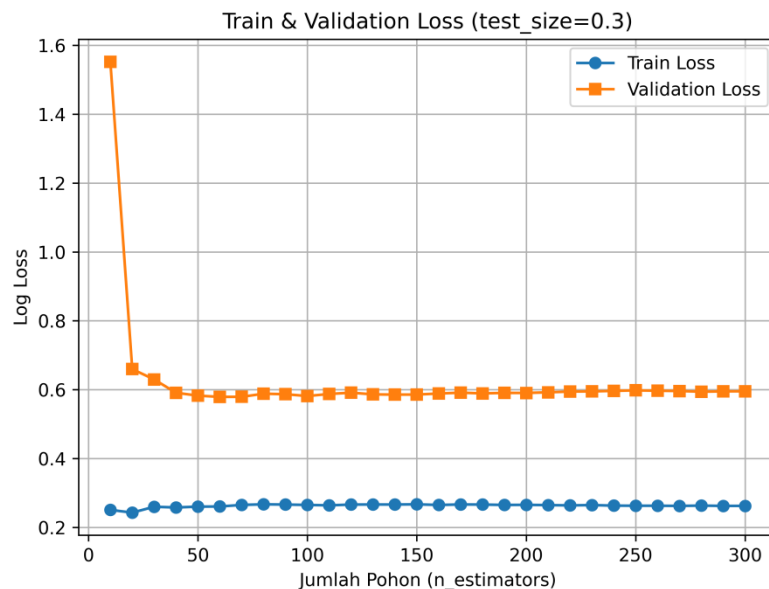
Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang digunakan untuk mengetahui performa model RF-B dalam mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 4.6.

Tabel 4. 6 Performa hasil pengujian model RF-B

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Baik	0.75	0.67	0.67	0.67	3
Baik Sekali		0.69	0.82	0.75	11
Unggul		0.88	0.7	0.78	10

4.2.3. Pelatihan Model RF-C

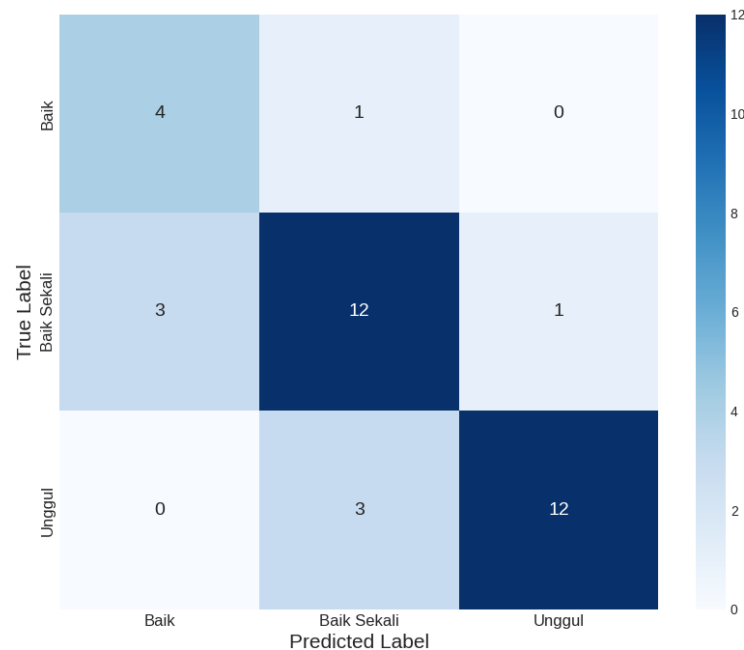
Pada proses eksperimen dari pelatihan model RF-C dalam klasifikasi kualitas perguruan tinggi di Indonesia digunakan proporsi data latih sebesar 70% atau sekitar 82 data dan data uji sebesar 30% atau sekitar 36 data. Model dilatih dengan jumlah pohon $n_estimators$ yang telah ditentukan, yaitu sebanyak 300 pohon. Hasil pelatihan menunjukkan nilai *train loss* sekitar 0.1432 dan nilai *validation loss* sekitar 0.5955 sebagaimana ditampilkan pada Gambar 4.7.



Gambar 4. 7 Grafik train dan validation loss model RF-C

Pada Gambar 4.7 menunjukkan nilai *train* dan *validation loss* pada proses pelatihan model RF-C yang terus mengalami penurunan hingga mencapai titik konvergensi pada $n_estimator$ ke-300, sehingga model tersebut dianggap telah optimal. Selanjutnya, model optimal yang telah didapatkan, kemudian digunakan untuk mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi dengan tiga (3) klasifikasi yaitu “Baik”, “Baik Sekali”, dan “Unggul”.

Hasil dari pengujian model RF-C dalam melakukan klasifikasi tingkat akreditasi perguruan tinggi disajikan pada Gambar 4.8.



Gambar 4. 8 Hasil klasifikasi model RF-C terhadap tingkat akreditasi

Hasil klasifikasi model RF-C terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 4.8 menunjukkan hasil klasifikasi model RF-C terhadap 36 data uji dalam tiga kategori kualitas perguruan tinggi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model berhasil mengklasifikasikan 4 dari 5 data dengan benar, meskipun masih terdapat 1 data yang salah diprediksi sebagai Baik Sekali. Pada kategori Baik Sekali, performa model terlihat dominan dengan 12 dari 16 data terklasifikasi benar, sementara 3 data salah diprediksi ke kategori Baik dan 1 data lainnya ke kategori Unggul. Adapun pada kategori Unggul, model mampu mengenali 12 dari 15 data dengan tepat, meskipun 3 data lainnya salah diprediksi ke kategori Baik Sekali.

Sintesisnya, hasil ini menunjukkan bahwa model RF-C cukup andal dalam mengenali ketiga kategori, meski masih terdapat tantangan dalam membedakan karakteristik yang berdekatan antar kelas, terutama antara Baik Sekali dan Unggul. Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-Score* maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 4.7.

Tabel 4. 7 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-C

Kategori	TP	TN	FP	FN
Baik	2	29	2	3
Baik Sekali	13	13	7	3
Unggul	11	20	1	4

Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang digunakan untuk mengetahui performa model RF-C dalam mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 4.8.

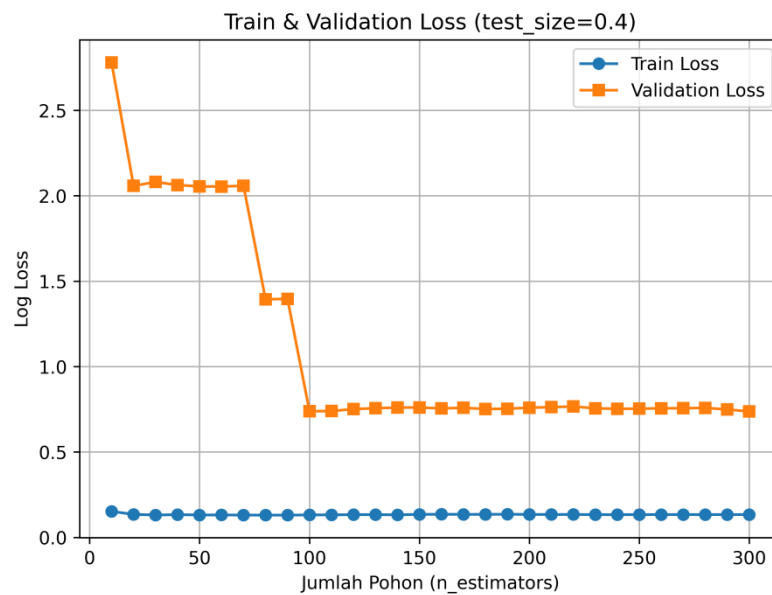
Tabel 4. 8 Performa hasil pengujian model RF-C

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Baik	0.78	0.57	0.8	0.67	5
Baik Sekali		0.75	0.75	0.75	16
Unggul		0.92	0.8	0.86	15

4.2.4. Pelatihan Model RF-D

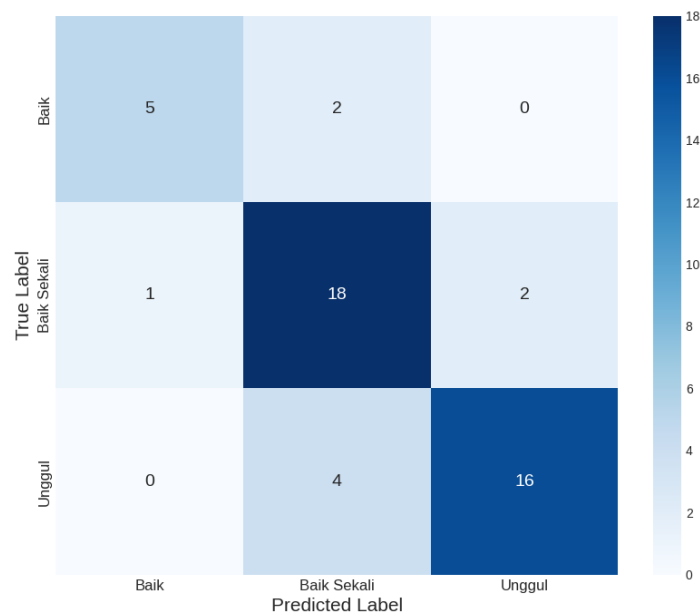
Pada proses eksperimen dari pelatihan model RF-D dalam klasifikasi kualitas perguruan tinggi di Indonesia digunakan proporsi data latih sebesar 60% atau sekitar 70 data dan data uji sebesar 40% atau sekitar 48 data. Model dilatih

dengan jumlah pohon keputusan ($n_estimators$) yang telah ditentukan, yaitu sebanyak 300 pohon. Hasil pelatihan menunjukkan nilai *train loss* sekitar 0.1335 dan nilai *validation loss* sekitar 0.7370 sebagaimana ditampilkan pada Gambar 4.9 di bawah ini:



Gambar 4. 9 Grafik train dan validation loss model RF-D

Pada Gambar 4.9 menunjukkan nilai *train loss* dan *validation loss* pada proses pelatihan model RF-D yang mengalami penurunan seiring dengan penambahan jumlah pohon hingga mencapai titik konvergensi pada sekitar $n_estimators$ ke-100. Setelah titik tersebut, nilai *validation loss* terlihat stabil meskipun jumlah pohon ditingkatkan hingga 300, sehingga model RF-D dianggap telah mencapai kondisi optimal. Model optimal yang diperoleh tersebut selanjutnya digunakan untuk melakukan klasifikasi kualitas perguruan tinggi di Indonesia berdasarkan tiga kategori akreditasi, yaitu “Baik”, “Baik Sekali”, dan “Unggul”. Hasil pengujian model RF-D dalam melakukan klasifikasi kualitas perguruan tinggi ditampilkan pada Gambar 4.9.



Gambar 4. 10 Hasil klasifikasi model RF-D terhadap tingkat akreditasi

Hasil klasifikasi model RF-D terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 4.10 menunjukkan hasil klasifikasi model RF-D terhadap 48 data uji dalam tiga kategori kualitas perguruan tinggi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model mampu mengklasifikasikan 5 dari 7 data dengan benar, meskipun masih terdapat 2 data yang salah diprediksi sebagai Baik Sekali. Pada kategori Baik Sekali, performa model terlihat dominan dengan 18 dari 21 data terklasifikasi tepat, sementara 1 data salah ke kategori Baik dan 2 data salah ke kategori Unggul. Adapun pada kategori Unggul, model berhasil mengenali 16 dari 20 data secara benar, meskipun 4 data lainnya salah diprediksi ke kategori Baik Sekali.

Secara keseluruhan, hasil ini menunjukkan bahwa model RF-D memiliki performa yang cukup baik dalam membedakan ketiga kategori, meskipun masih

terdapat tantangan dalam memisahkan karakteristik yang berdekatan, terutama antara Baik Sekali dan Unggul. Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-Score* maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 4.9.

Tabel 4. 9 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-D

Kategori	TP	TN	FP	FN
Baik	5	40	1	2
Baik Sekali	18	21	6	3
Unggul	16	26	2	4

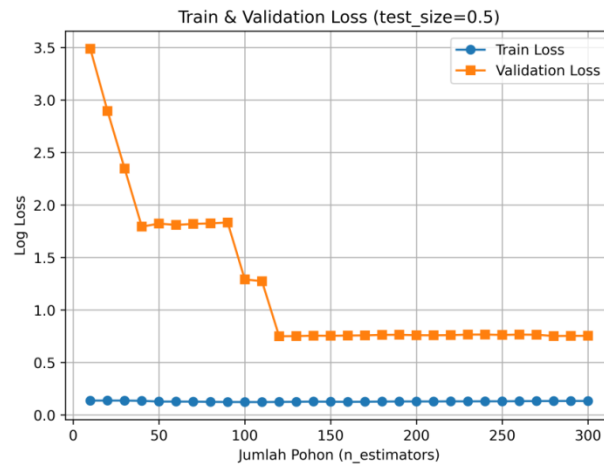
Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang digunakan untuk mengetahui performa model RF-D dalam mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 4.10.

Tabel 4. 10 Performa hasil pengujian model RF-D

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Baik	0.81	0.83	0.71	0.77	7
Baik Sekali		0.75	0.86	0.8	21
Unggul		0.89	0.8	0.84	20

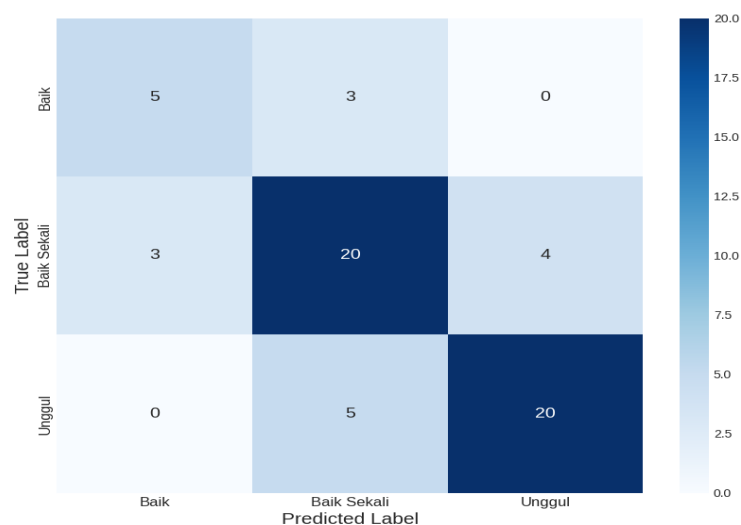
4.2.5. Pelatihan Model RF-E

Pada proses eksperimen dari pelatihan model RF-E dalam klasifikasi kualitas perguruan tinggi di Indonesia digunakan proporsi data latih sebesar 50% atau sekitar 58 data dan data uji sebesar 50% atau sekitar 60 data. Model dilatih dengan jumlah *n_estimators* yang telah ditentukan, yaitu sebanyak 300 pohon. Hasil pelatihan menunjukkan nilai *train loss* sekitar 0.1333 dan nilai *validation loss* sekitar 0.7540 sebagaimana ditampilkan pada Gambar 4.11.



Gambar 4. 11 Grafik train dan validation loss model RF-E

Pada Gambar 4.11 menunjukkan nilai *train* dan *validation loss* pada proses pelatihan model RF-E yang terus mengalami penurunan hingga mencapai titik konvergensi pada $n_{estimator}$ ke-300, sehingga model tersebut dianggap telah optimal. Selanjutnya, model optimal yang telah didapatkan, kemudian digunakan untuk mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi dengan tiga (3) klasifikasi yaitu “Baik”, “Baik Sekali”, dan “Unggul”. Hasil dari pengujian model RF-E dalam melakukan klasifikasi tingkat akreditasi perguruan tinggi disajikan pada Gambar 4.12.



Gambar 4. 12 Hasil klasifikasi model RF-E terhadap tingkat akreditasi

Hasil klasifikasi model RF-E terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 4.12 memperlihatkan hasil klasifikasi model RF-E terhadap 60 data uji dalam tiga kategori kualitas perguruan tinggi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model berhasil mengklasifikasikan 5 dari 8 data dengan benar, sementara 3 data lainnya salah diprediksi sebagai Baik Sekali. Pada kategori Baik Sekali, performa model cukup dominan dengan 20 dari 27 data terklasifikasi tepat, meskipun masih terdapat 3 data yang salah ke kategori Baik dan 4 data lainnya ke kategori Unggul. Adapun pada kategori Unggul, model mampu mengenali 20 dari 25 data dengan benar, sedangkan 5 data lainnya salah diprediksi ke kategori Baik Sekali.

Kesimpulannya, hasil tersebut menunjukkan bahwa model RF-E cukup andal dalam mengklasifikasikan ketiga kategori, meskipun masih terdapat tantangan dalam membedakan karakteristik antar kelas yang berdekatan, khususnya antara Baik Sekali dan Unggul. Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-Score* maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 4.11.

Tabel 4. 11 Nilai TP, TN, FP, dan FN setiap kategori pada model RF-E

Kategori	TP	TN	FP	FN
Baik	4	48	3	4
Baik Sekali	19	24	9	7
Unggul	20	30	4	5

Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang

digunakan untuk mengetahui performa model RF-E dalam mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 4.12.

Tabel 4. 12 Performa hasil pengujian model RF-E

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Baik	0.75	0.62	0.62	0.62	8
Baik Sekali		0.71	0.74	0.73	27
Unggul		0.83	0.8	0.82	25

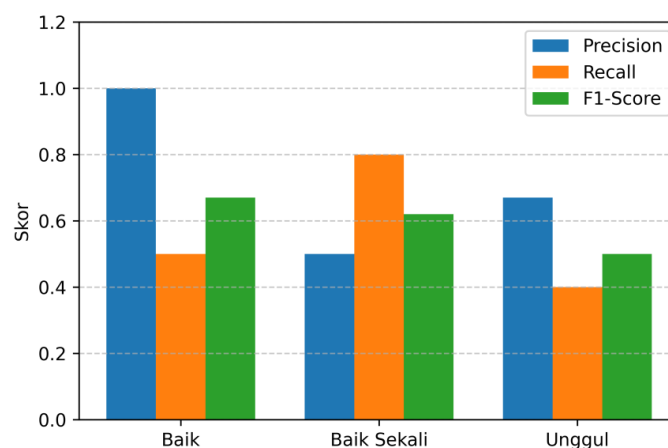
4.3. Ujicoba Model *Random Forest*

Setelah proses pelatihan selesai, masing-masing model *Random Forest* ujicoba menggunakan data uji guna menilai kinerjanya dalam mengklasifikasikan kualitas perguruan tinggi ke dalam tiga kategori: Baik, Baik Sekali, dan Unggul. Evaluasi ini menggunakan sejumlah metrik penting, yakni *accuracy*, *precision*, *recall*, dan *F1-Score*, yang masing-masing memberikan perspektif berbeda terhadap performa model.

Tabel 4. 13 Hasil ujicoba klasifikasi kualitas model *Random Forest*

Model	Kategori	Accuracy	Precision	Recall	F1-Score	Support
RF-A	Baik	0.58	1	0.5	0.67	2
	Baik Sekali		0.5	0.8	0.62	5
	Unggul		0.67	0.4	0.5	5
RF-B	Baik	0.75	0.67	0.67	0.67	3
	Baik Sekali		0.69	0.82	0.75	11
	Unggul		0.88	0.7	0.78	10
RF-C	Baik	0.78	0.57	0.8	0.67	5
	Baik Sekali		0.75	0.75	0.75	16
	Unggul		0.92	0.8	0.86	15
RF-D	Baik	0.81	0.83	0.71	0.77	7
	Baik Sekali		0.75	0.86	0.8	21
	Unggul		0.89	0.8	0.84	20
RF-E	Baik	0.75	0.62	0.62	0.62	8
	Baik Sekali		0.71	0.74	0.73	27
	Unggul		0.83	0.8	0.82	25

Pada Tabel 4.13 menunjukkan bahwa model RF-A menghasilkan nilai accuracy sebesar 0.58, yang berarti sekitar 58.33% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Untuk kategori Baik, precision mencapai 1.00 atau sekitar 100%, namun recall hanya 0.50 atau sekitar 50%, sehingga meskipun prediksi benar sangat akurat, model masih gagal mengenali seluruh data aktual. Pada kategori Baik Sekali, precision rendah (0.50 atau sekitar 50%) tetapi recall tinggi (0.80 atau sekitar 80%), menunjukkan sensitivitas yang baik meski ketepatan masih lemah. Sementara itu, kategori Unggul memiliki precision 0.67 atau sekitar 67% dan recall 0.40 atau sekitar 40%, yang menandakan kinerja kurang optimal. Nilai f1-score keseluruhan (0.67 atau sekitar 67%; 0.62 atau sekitar 62%; 0.50 atau sekitar 50%) mencerminkan bahwa model cukup baik untuk kelas tertentu, namun performanya belum stabil karena jumlah data uji yang terbatas.

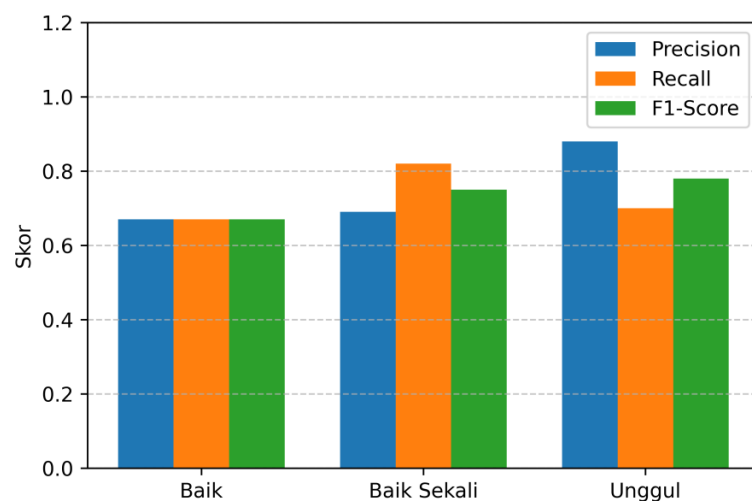


Gambar 4. 13 Perbandingan matrik evaluasi model RF-A

Secara keseluruhan, performa model RF-A pada Gambar 4.13 menunjukkan bahwa meskipun model memiliki ketepatan prediksi yang cukup baik, kemampuannya dalam mendeteksi semua kategori secara merata masih perlu

ditingkatkan. Hal ini penting agar model tidak hanya “benar saat menebak”, tetapi juga mampu mengenali semua data aktual dengan adil dan menyeluruh. Temuan ini menjadi dasar pertimbangan untuk melakukan perbaikan model melalui teknik seperti penyeimbangan data, tuning parameter, maupun eksplorasi algoritma lain yang mungkin lebih cocok untuk karakteristik data yang digunakan.

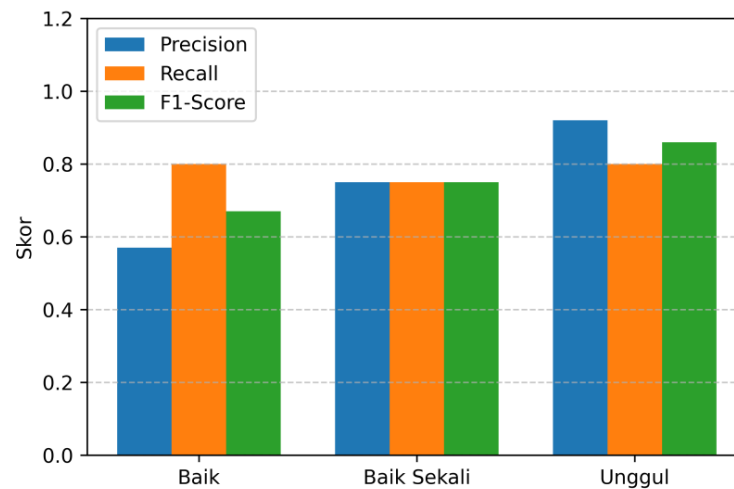
Tabel 4.13 menunjukkan bahwa model RF-B menghasilkan nilai accuracy sebesar 0.75, yang berarti sekitar 75% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Pada kategori Kategori Baik memperoleh nilai precision, recall, dan f1-score sebesar 0.67 atau sekitar 67%, namun hasil ini kurang stabil karena hanya didukung 3 data. Kategori Baik Sekali menunjukkan performa lebih seimbang dengan precision 0.69 atau sekitar 69%, recall 0.82 atau sekitar 82%, dan f1-score 0.75 atau sekitar 75% dari 11 data, sehingga lebih representatif. Sementara itu, kategori Unggul menonjol dengan precision 0.88 atau sekitar 88% dan f1-score 0.78 atau sekitar 78%, meskipun recall lebih rendah pada 0.70 atau sekitar 70%.



Gambar 4. 14 Perbandingan matrik evaluasi model RF-B

Secara keseluruhan, evaluasi terhadap model RF-B pada Gambar 4.14 menunjukkan performa yang lebih baik dibandingkan RF-A. Model ini memiliki tingkat akurasi dan keseimbangan klasifikasi yang cukup tinggi di antara semua kategori. Hal ini menunjukkan bahwa pendekatan yang diterapkan dalam RF-B baik dari segi konfigurasi data latih maupun parameter model lebih mampu menangkap pola umum dalam data dan melakukan klasifikasi secara adil. Meskipun demikian, pengujian lanjutan dan tuning lebih lanjut tetap dapat dilakukan untuk mengoptimalkan performa model ke tingkat yang lebih tinggi.

Kemudian, pada Tabel 4.13 menunjukkan bahwa model RF-C menghasilkan nilai akurasi sebesar 0.78, yang berarti sekitar 77.78% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Untuk kategori Baik memperoleh nilai precision 0.57 atau sekitar 57%, recall 0.80 atau sekitar 80%, dan f1-score 0.67 atau sekitar 67%. Hasil ini menunjukkan kemampuan model yang cukup baik dalam menangkap data aktual dengan jumlah data uji kecil (5 data). Kategori Baik Sekali tampil lebih konsisten dengan precision, recall, dan f1-score sama-sama 0.75 atau sekitar 75%, didukung oleh jumlah data lebih banyak (16 data) sehingga lebih dapat diandalkan. Sementara itu, kategori Unggul menonjol dengan precision 0.92 atau sekitar 92%, recall 0.80 atau sekitar 80%, dan f1-score 0.86 atau sekitar 86%, menunjukkan kinerja paling efektif dalam identifikasi kelas ini, terlebih dengan dukungan data yang cukup besar (15 data).

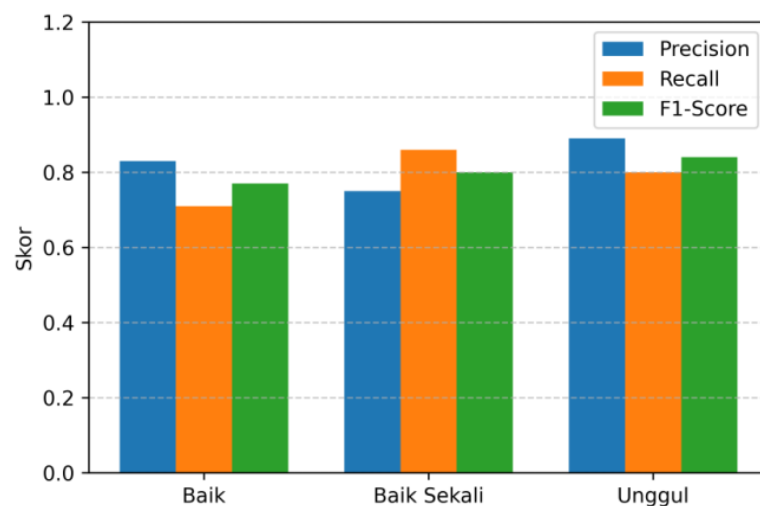


Gambar 4. 15 Perbandingan matrik evaluasi model RF-C

Secara keseluruhan, hasil evaluasi terhadap model RF-C pada Gambar 4.15 menunjukkan peningkatan performa dibandingkan model-model sebelumnya. Model ini tidak hanya unggul dalam akurasi keseluruhan, tetapi juga menunjukkan keseimbangan dalam mengenali dan mengklasifikasikan seluruh kelas secara adil. Dengan hasil ini, RF-C layak dijadikan kandidat utama dalam sistem klasifikasi, meskipun pengujian lanjutan dan penyempurnaan melalui tuning parameter atau strategi pengolahan data tetap disarankan untuk mencapai performa maksimal.

Selanjutnya, pada Tabel 4.13 menunjukkan bahwa model RF-D menghasilkan nilai accuracy sebesar 0.81, yang berarti sekitar 81.25% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Untuk kategori Baik memperoleh nilai precision 0.83 atau sekitar 83%, recall 0.71 atau sekitar 71%, dan f1-score 0.77 atau sekitar 77%. Hasil ini menunjukkan keseimbangan cukup baik antara ketepatan dan kelengkapan, meskipun dengan dukungan data yang relatif sedikit (7 data), sehingga interpretasi tetap perlu hati-

hati. Kategori Baik Sekali mencatat precision 0.75 atau sekitar 75%, recall 0.86 atau sekitar 86%, dan f1-score 0.80 atau sekitar 80%, yang menandakan performa seimbang dan cukup andal. Dukungan data yang lebih besar (21 data) membuat hasil kategori ini lebih stabil dan dapat dipercaya. Sementara itu, kategori Unggul menampilkan performa terbaik dengan precision 0.89 atau sekitar 89%, recall 0.80 atau sekitar 80%, dan f1-score 0.84 atau sekitar 84%, yang mencerminkan efektivitas tinggi model RF-D dalam mengklasifikasikan data (20 data).

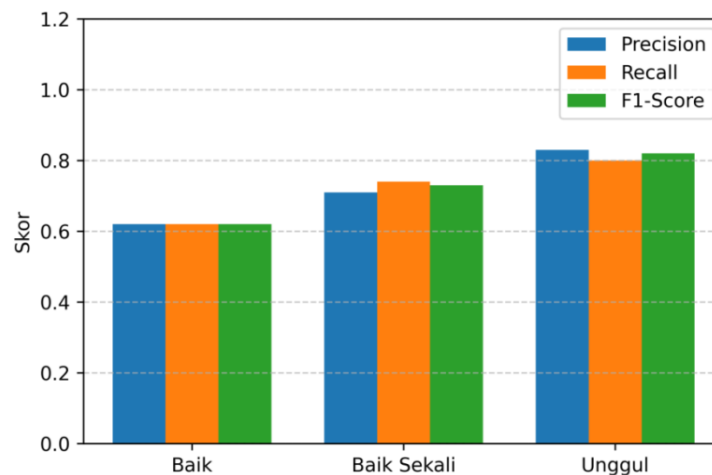


Gambar 4. 16 Perbandingan matrik evaluasi model RF-D

Secara keseluruhan, evaluasi terhadap model RF-D pada Gambar 4.16 menunjukkan peningkatan performa yang signifikan dibandingkan model-model sebelumnya, baik dari sisi akurasi maupun keseimbangan antar kelas. Dengan tingkat f1-score yang tinggi dan stabilitas dalam precision dan recall, model ini layak dipertimbangkan sebagai solusi klasifikasi yang andal. Meski demikian, optimalisasi lanjutan melalui tuning parameter, validasi silang, atau pemanfaatan fitur tambahan tetap dianjurkan untuk mencapai performa maksimum sesuai konteks aplikasinya.

Selanjutnya, pada Tabel 4.13 menunjukkan bahwa model RF-E menghasilkan nilai akurasi sebesar 0.75, yang berarti sekitar 75% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Untuk kategori Baik memperoleh nilai precision 0.62 atau sekitar 62%, recall 0.62 atau sekitar 62%, dan f1-score 0.62 atau sekitar 62%. Performa ini tergolong moderat, terlebih dengan dukungan data yang terbatas (8 data), sehingga hasilnya masih rentan fluktuasi dan menunjukkan ruang perbaikan. Kategori Baik Sekali tampil lebih stabil dengan precision 0.71 atau sekitar 71%, recall 0.74 atau sekitar 74%, dan f1-score 0.73 atau sekitar 73%, yang mencerminkan keseimbangan baik antara ketepatan dan cakupan. Dukungan data yang lebih besar (27 data) membuat hasil kategori ini lebih kuat dan dapat diandalkan. Sementara itu, kategori Unggul menunjukkan performa terbaik dengan precision 0.83 atau sekitar 83%, recall 0.80 atau sekitar 80%, dan f1-score 0.82 atau sekitar 82%, menandakan sangat efektif mengenali kategori ini dengan dukungan data uji yang cukup besar (25 data).

Secara keseluruhan, evaluasi terhadap model RF-E pada Gambar 6.5 menunjukkan performa klasifikasi yang tergolong baik dan stabil, meskipun belum sekuat model RF sebelumnya seperti RF-D. Model RF-E dapat dijadikan alternatif dalam skenario klasifikasi, khususnya ketika akurasi dan keseimbangan antar kelas tidak menjadi syarat utama. Namun, untuk mencapai performa optimal, disarankan untuk melakukan perbaikan lanjutan seperti penyesuaian parameter model, penyeimbangan distribusi kelas, atau pengayaan fitur yang lebih relevan terhadap klasifikasi.



Gambar 4. 17 Perbandingan matrik evaluasi model RF-E

Berdasarkan hasil ujicoba terhadap klasifikasi kualitas perguruan tinggi pada lima skenario eksperimen model *Random Forest* (RF-A hingga RF-E), terlihat bahwa variasi rasio data latih dan uji memberikan pengaruh yang signifikan terhadap performa klasifikasi model. Setiap skenario memiliki keunggulan dan kelemahan tersendiri dalam proses klasifikasi perguruan tinggi ke dalam tiga kelas utama: Baik, Baik Sekali, dan Unggul.

Tabel 4. 14 Performa model *Random Forest*

Nama Pelatihan	Rasio Data	Nilai Akurasi
RF-A	90:10	58.33%
RF-B	80:20	75%
RF-C	70:30	77.78%
RF-D	60:40	81.25%
RF-E	50:50	75%

Pada Tabel 4.14 menunjukkan bahwa model RF-A, yang menggunakan rasio data latih dan uji sebesar 90:10, mencatatkan akurasi sebesar 58,33%, yang merupakan akurasi terendah di antara kelima skenario. Meskipun porsi data latih sangat besar, jumlah data uji yang sangat sedikit kemungkinan besar menyebabkan model overfitting dan hasil evaluasi menjadi kurang representatif.

Selanjutnya, model RF-B dengan rasio 80:20 mengalami peningkatan performa yang signifikan, dengan nilai akurasi mencapai 75,00%. Peningkatan ini menunjukkan bahwa penambahan porsi data uji memberikan informasi evaluasi yang lebih stabil, meskipun masih terdapat tantangan dalam mengklasifikasikan kategori tertentu secara konsisten.

Model RF-C, yang menerapkan rasio data latih dan uji sebesar 70:30, menunjukkan performa terbaik dengan akurasi tertinggi mencapai 77,78%. Rasio ini memberikan keseimbangan optimal antara proses pelatihan dan pengujian, sehingga model mampu mengenali pola data dengan lebih akurat dan menyeluruh. Sementara itu, model RF-D dengan rasio 60:40 mencatatkan akurasi sebesar 81,25%, yang justru lebih tinggi daripada RF-C. Namun demikian, nilai ini perlu dianalisis lebih lanjut dari segi stabilitas klasifikasi dan distribusi prediksi antarkategori, karena peningkatan data uji yang terlalu besar berisiko menurunkan kemampuan pembelajaran model.

Terakhir, model RF-E dengan rasio data yang seimbang (50:50) menunjukkan akurasi sebesar 75,00%. Meski tidak menjadi yang tertinggi, model ini memberikan distribusi klasifikasi yang cukup proporsional dan cenderung adil antarkelas. Rasio yang seimbang ini juga memberi gambaran realistis terkait kemampuan model dalam kondisi data nyata, meskipun risiko underfitting tetap harus diperhatikan.

Secara keseluruhan, performa terbaik secara akurasi ditunjukkan oleh model **RF-D** dengan nilai akurasi sekitar 81,25%, diikuti oleh RF-C dengan nilai akurasi

sekitar 77,78%, dan RF-B serta RF-E dengan nilai masing-masing akurasi sekitar 75%. Sementara itu, RF-A menjadi model dengan performa terendah dengan akurasi sekitar 58,33%. Temuan ini menunjukkan pentingnya pemilihan rasio data latih dan uji yang tepat dalam membangun model klasifikasi yang andal untuk penilaian mutu perguruan tinggi. Selain itu, penting pula untuk mempertimbangkan evaluasi per kategori dan kecenderungan bias klasifikasi, seperti middle bias, guna menghasilkan sistem prediksi yang lebih akurat dan adil.

Tabel 4. 15 Nilai TP, TN, FP, FN model RF-D

Kelas	True Positive	True Negative	False Positive	False Negative
Baik	5	40	1	2
Baik Sekali	18	21	6	3
Unggul	16	26	2	4

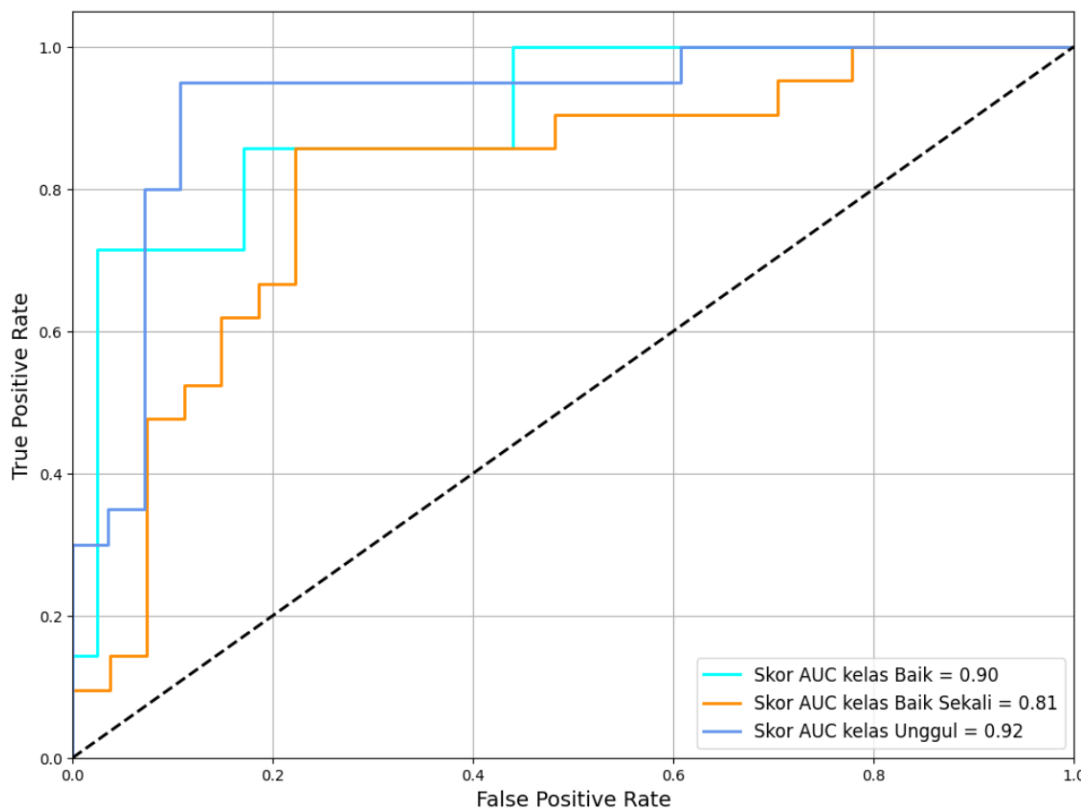
Tabel 4.15 menyajikan rincian performa model RF-D dalam klasifikasi perguruan tinggi ke dalam tiga kategori, yaitu Baik, Baik Sekali, dan Unggul, berdasarkan metrik True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Nilai-nilai ini memberikan gambaran yang lebih mendalam mengenai sejauh mana model mampu membedakan antar kelas serta kecermatannya dalam melakukan prediksi.

Pada kategori Baik, model mampu mengidentifikasi 5 data secara benar (TP) dan 40 data lainnya secara akurat sebagai bukan Baik (TN). Namun, terdapat 2 data yang seharusnya termasuk kategori Baik tetapi salah diklasifikasikan sebagai kelas lain (FN), serta 1 data yang keliru diklasifikasikan sebagai Baik padahal bukan (FP). Hal ini menunjukkan bahwa meskipun model cukup baik dalam mengenali data non-Baik, masih terdapat kelemahan dalam mendeteksi seluruh data yang benar-benar termasuk dalam kategori tersebut. Rendahnya

jumlah False Positive mengindikasikan bahwa model cukup hati-hati dalam menetapkan label Baik, namun beberapa data terlewatkan, yang menurunkan nilai recall pada kategori ini.

Sementara itu, pada kategori Baik Sekali, model mencatat 18 data yang berhasil diklasifikasikan dengan benar (TP) dan 21 data lainnya dikenali secara tepat sebagai bukan bagian dari kelas ini (TN). Namun, terdapat 6 data yang salah diklasifikasikan sebagai Baik Sekali (FP) dan 3 data yang seharusnya masuk kategori ini namun terlewat (FN). Jumlah False Positive yang tinggi dibandingkan kategori lain menunjukkan bahwa model cenderung lebih sering memilih kelas ini saat prediksi kurang jelas. Fenomena ini dapat mencerminkan adanya middle bias, yaitu kecenderungan model untuk memilih kelas menengah saat menghadapi ambiguitas, mengingat Baik Sekali berada di tengah-tengah antara Baik dan Unggul.

Pada kategori Unggul, performa model menunjukkan hasil yang cukup stabil. Terdapat 16 data yang berhasil dikenali secara akurat sebagai Unggul (TP) dan 26 data yang tepat diklasifikasikan sebagai bukan Unggul (TN). Meski demikian, model masih melakukan kesalahan dengan mengklasifikasikan 2 data sebagai Unggul padahal bukan (FP), serta gagal mengenali 4 data yang seharusnya termasuk dalam kategori ini (FN). Meskipun terdapat kekeliruan, jumlah kesalahan pada kategori ini relatif lebih rendah dibandingkan kategori lain, yang mencerminkan kemampuan model yang lebih andal dalam mendeteksi mutu tertinggi perguruan tinggi.



Gambar 4. 18 Skor AUC model RF-D

Untuk dapat memahami seberapa baik model dapat membedakan satu kelas dengan kelas lainnya, dalam penelitian ini dilakukan analisis lanjutan pada model yang memiliki performa tinggi menggunakan analisis Kurva ROC (Receiver Operating Characteristic). Hasil Visualisasi Kurva ROC dari model **RF-D** dalam penelitian ini yang ditampilkan pada Gambar 4.18 memperoleh skor AUC (Area Under Curve) untuk kelas ‘Unggul’ sebesar 0.92, kelas ‘Baik Sekali’ sebesar 0.81, dan untuk kelas ‘Baik’ sebesar 0.90. Skor AUC yang mendekati 1.0 ini mengindikasikan bahwa model memiliki performa yang tinggi untuk dapat membedakan antara perguruan tinggi di Indonesia yang memiliki akreditasi ‘Unggul’, ‘Baik Sekali’, dan ‘Baik’.

4.4. Kesimpulan

Berdasarkan hasil evaluasi menyeluruh terhadap model *Random Forest* (RF) dalam mengukur kualitas perguruan tinggi di Indonesia, dapat disimpulkan bahwa performa model sangat dipengaruhi oleh komposisi data latih dan uji yang digunakan. Dari kelima skenario yang diuji (RF-A hingga RF-E), model RF-D menunjukkan performa paling optimal dengan akurasi sebesar 81,25%, diikuti oleh RF-C (77,78%), serta RF-B dan RF-E (75%). Sementara itu, RF-A menghasilkan akurasi terendah (58,33%), menunjukkan bahwa penggunaan data uji yang terlalu kecil (10%) dapat menyebabkan overfitting dan hasil yang kurang representatif.

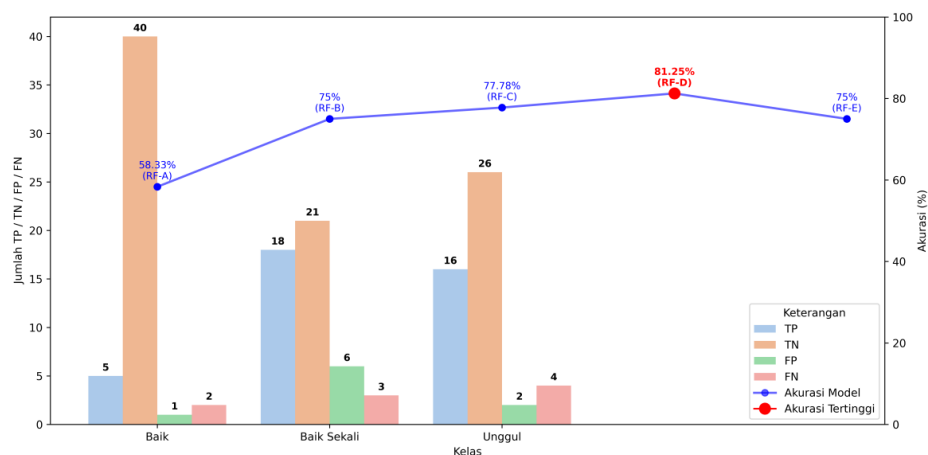
Evaluasi berdasarkan precision, recall, dan f1-score menunjukkan bahwa model RF-D tidak hanya unggul dalam akurasi, tetapi juga mampu menjaga keseimbangan performa antar kategori. Kelas “Unggul” memperoleh nilai f1-score yang tinggi, menunjukkan kemampuan model dalam mendeteksi kategori mutu tertinggi dengan baik. Namun, middle bias masih ditemukan, khususnya pada kelas “Baik Sekali”, yang terlihat dari peningkatan jumlah False Positive (FP) dan False Negative (FN) menunjukkan kecenderungan model untuk memilih kategori tengah saat menghadapi ambiguitas.

Analisis lanjutan terhadap nilai TP, TN, FP, FN pada model RF-D memperkuat temuan tersebut. Meskipun nilai True Negative (TN) relatif tinggi untuk seluruh kelas, masih terdapat tantangan dalam meminimalkan False Negative, terutama pada kategori “Baik”. Hal ini berdampak pada penurunan

recall dan menunjukkan bahwa model kadang gagal mengenali seluruh data yang seharusnya termasuk dalam kategori tertentu.

Untuk mengukur kemampuan diskriminatif model dalam membedakan antar kelas, dilakukan analisis ROC Curve dan perhitungan AUC (Area Under Curve). Model RF-D menunjukkan hasil yang sangat baik, dengan skor AUC sebesar 0.92 untuk kelas “Unggul”, 0.90 untuk kelas “Baik”, dan 0.81 untuk kelas “Baik Sekali”. Skor AUC yang tinggi ini mengindikasikan bahwa model memiliki kemampuan klasifikasi yang kuat dan stabil, khususnya dalam membedakan kualitas perguruan tinggi di Indonesia secara menyeluruh.

Temuan penelitian ini menunjukkan bahwa model *Random Forest* (RF), khususnya pada skenario **RF-D**, memiliki performa paling optimal dalam mengukur kualitas perguruan tinggi, baik dari segi akurasi, keseimbangan klasifikasi, hingga nilai AUC. Hasil ini sejalan dengan penelitian oleh Mohan *et al.*, (2024) yang menyatakan bahwa *Random Forest* unggul dalam menangani data klasifikasi multikategori karena kemampuannya mengelola hubungan non-linier dan noise dalam data pendidikan.



Gambar 4. 19 Perbandingan performa model *Random Forest*

Penelitian oleh Bhuvaneswari & Karthikeyan, (2023) juga mendukung temuan ini, di mana *Random Forest* terbukti lebih unggul dibanding *Logistic Regression* dalam tugas prediksi mutu akademik karena memberikan akurasi dan generalisasi yang lebih baik. Sebaliknya, performa *Logistic Regression* dalam penelitian ini cenderung stabil namun tidak seakurat *Random Forest*. Hal ini juga diamini oleh Al-Batah *et al.*, (2024) yang menyatakan bahwa *Logistic Regression* lebih tepat digunakan untuk klasifikasi biner atau masalah dengan relasi linier yang jelas antar variabel. Dalam konteks klasifikasi kualitas perguruan tinggi yang bersifat kompleks dan tidak selalu linier, model seperti *Random Forest* terbukti lebih adaptif dan akurat. Temuan ini memperkuat bahwa dalam klasifikasi mutu pendidikan tinggi, pemilihan algoritma harus mempertimbangkan kompleksitas data dan struktur kategorinya.

BAB V

KLASIFIKASI KUALITAS PERGURUAN TINGGI DENGAN *LOGISTIC REGRESSION*

Logistic Regression adalah suatu metode yang dapat digunakan untuk mencari hubungan antara variabel respon yang bersifat *Dichotomous* (skala nominal/ordinal dengan dua kategori) atau *Polychotomous* (skala nominal/ordinal dengan lebih dari dua kategori) dengan satu atau lebih variabel prediktor berskala kategori atau kontinyu. Model *Logistic Regression* terdiri atas *Logistic Regression* dengan respons biner, ordinal, dan multinomial. *Logistic Regression* biner adalah suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon (y) yang bersifat biner (*dichotomous*) dengan variabel prediktor (x) yang bersifat kategorik atau kontinyu (Hosmer *et al.*, 2013).

Hasil dari respon variabel *Dichotomous* memiliki dua kategori, yaitu “sukses/gagal”, “ada/tidak ada” dan sebagainya yang di jelaskan dalam $y = 1$ mewakili kemungkinan sukses dengan probabilitas $\pi(x)$; $y = 0$ mewakili kemungkinan gagal dengan probabilitas $1 - \pi(x)$. Dimana setiap pengamatan mengikuti proses bernouli (Hosmer *et al.*, 2013).

Pada *Logistic Regression* dapat disusun model yang terdiri dari banyak variabel prediktor dikenal sebagai Model Regresi Logistik Multivariabel. Rata-rata bersyarat dari y jika diberikan nilai x di tuliskan dalam bentuk $(x) = (y|x)$. Pernyataan persamaan Model *Logistic Regression* Multivariabel dengan p variabel prediktor adalah sebagai berikut (Hosmer *et al.*, 2013) :

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon}} \quad (5.1)$$

Keterangan:

- $\pi(x)$: Probabilitas kejadian dari kelas positif.
- e : Bilangan Euler (sekitar 2.71828).
- β_0 : Intersep (bias).
- $\beta_1, \beta_2, \dots, \beta_p$: Koefisien regresi untuk tiap variabel prediktor.
- x_1, x_2, \dots, x_p : Nilai fitur/prediktor.
- ε : *Error* (kadang tidak di tulis secara eksplisit karena *Logistik Regression* biasanya memodelkan *error* dalam probabilitas).

Fungsi ini akan menghasilkan *output* di antara 0 dan 1, yang kemudian bisa diinterpretasikan sebagai probabilitas untuk klasifikasi biner. Menurut Hosmer at al. (2013), bahwa untuk mempermudah pendugaan parameter regresi di gunakan transformasi *logit* dari (x) yang dirumuskan pada persamaan 3.6 sebagai berikut :

$$\begin{aligned} g(x) &= \ln \left[\frac{\pi(x)}{1-\pi(x)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon \\ &= \sum_{k=0}^p \beta_k x_k \\ &= x^T \beta \end{aligned} \quad (5.2)$$

Dari persamaan pada 5.2 dijabarkan sebagai berikut:

1) Persamaan awal

Berdasarkan persamaan 5.2 dapat kita jelaskan secara rinci bahwa pada bagian persamaan berikut :

$$g(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right),$$

disebut dengan model logit. Persamaan ini menunjukkan bahwa rasio peluang atau *odds* ($\pi(x)/(1 - \pi(x))$) menggambarkan peluang keberhasilan dibandingkan dengan kegagalan. Berikut merupakan keterangan dari rumus model logit yaitu :

- $\pi(x)$ = probabilitas kejadian sukses (kelas positif)
- $1-\pi(x)$ = probabilitas kejadian gagal (kelas negatif)
- Jadi, $\pi(x)/1-\pi(x)$ adalah **odds** peluang keberhasilan dibandingkan kegagalan.

2) Hasilnya

Dari persamaan 5.2 berdasarkan persamaan awal, maka akan menghasilkan persamaan berikut :

$$= \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon \quad (5.3)$$

Persamaan 5.3 adalah *linear predictor* atau komponen linear dari model sama seperti pada regresi linear. Namun pada *Logistic Regression*, komponen linear ini tidak langsung menjadi output, melainkan menjadi input untuk fungsi logit, sehingga output-nya adalah *log-odds*.

3) Notasi ringkas

1. *Sigma Notation*

$$= \sum_{k=0}^p \beta_k x_k \quad (5.4)$$

- $k = 0$ untuk intercept β_0 , dan $x_0 = 1$.
- $k = 1 \dots p$ untuk semua fitur.

2. *Vektor-Matriks Notation*

$$= x^T \beta \quad (5.5)$$

- x adalah vektor fitur.
- β adalah vektor koefisien.
- $x^T \beta$ adalah hasil dot product antara fitur dan koefisiennya.

Intinya persamaan ini menjelaskan hubungan linear antara fitur-fitur x dan *log-odds* probabilitas kejadian. Walaupun modelnya linear dalam *log-odds*, hasil akhirnya tetap dalam bentuk probabilitas (karena di bungkus lagi dengan fungsi sigmoid/*logistic*).

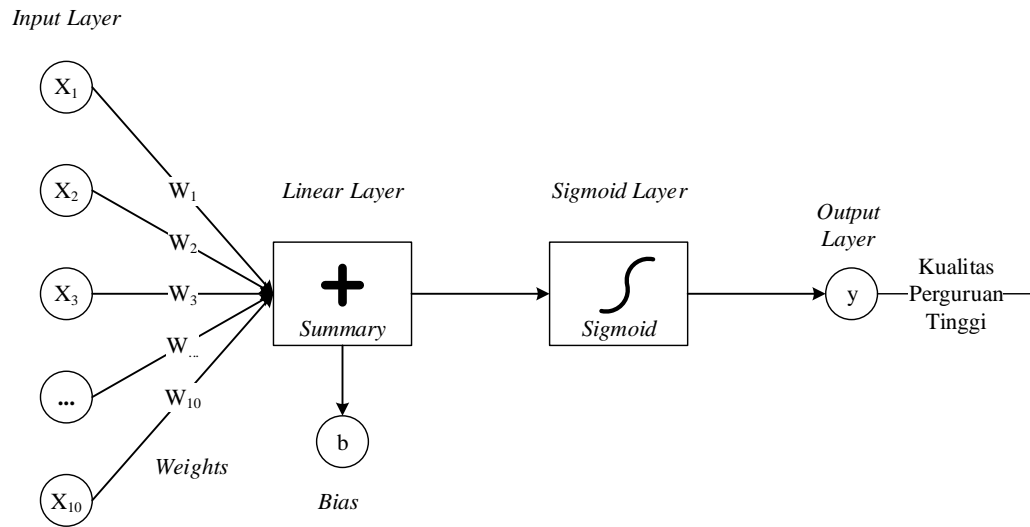
Selanjutnya model *Logistic Regression* pada persamaan (5.2) dapat dituliskan dalam bentuk:

$$\pi(x) = \frac{\exp(g(x))}{1+\exp(g(x))} \quad (5.6)$$

Persamaan 5.6 merupakan hasil kebalikan (inversi) dari transformasi logit, yang mengubah nilai linear $g(x)$ menjadi probabilitas $\pi(x)$ dalam rentang 0 hingga 1. Bentuk fungsi ini dikenal sebagai Fungsi Logistik (*Sigmoid Function*) dan merupakan dasar utama dari Model *Logistic Regression*. (Hosmer *et al.*, 2013)

5.1. Desain Model *Logistic Regression*

Desain penelitian ini menjelaskan tahapan-tahapan yang ditempuh peneliti untuk memperoleh hasil yang ditujukan sebagai jawaban atas permasalahan yang diangkat. Pada penelitian ini digunakan model *Logistic Regression* (LR) dengan skema pelatihan mulai dari LR-A hingga LR-E untuk melakukan klasifikasi tingkat akreditasi dalam rangka mengukur kualitas perguruan tinggi di Indonesia. Model *Logistic Regression* dipilih karena kemampuannya dalam mengolah data dengan variabel prediktor yang beragam serta menghasilkan probabilitas klasifikasi yang jelas untuk setiap kategori akreditasi. Dengan pendekatan ini, diharapkan model mampu memberikan hasil yang sederhana namun tetap akurat dalam memprediksi kualitas perguruan tinggi. Gambaran lebih rinci mengenai rancangan arsitektur model *Logistic Regression* ini dapat dilihat pada Gambar 5.1.

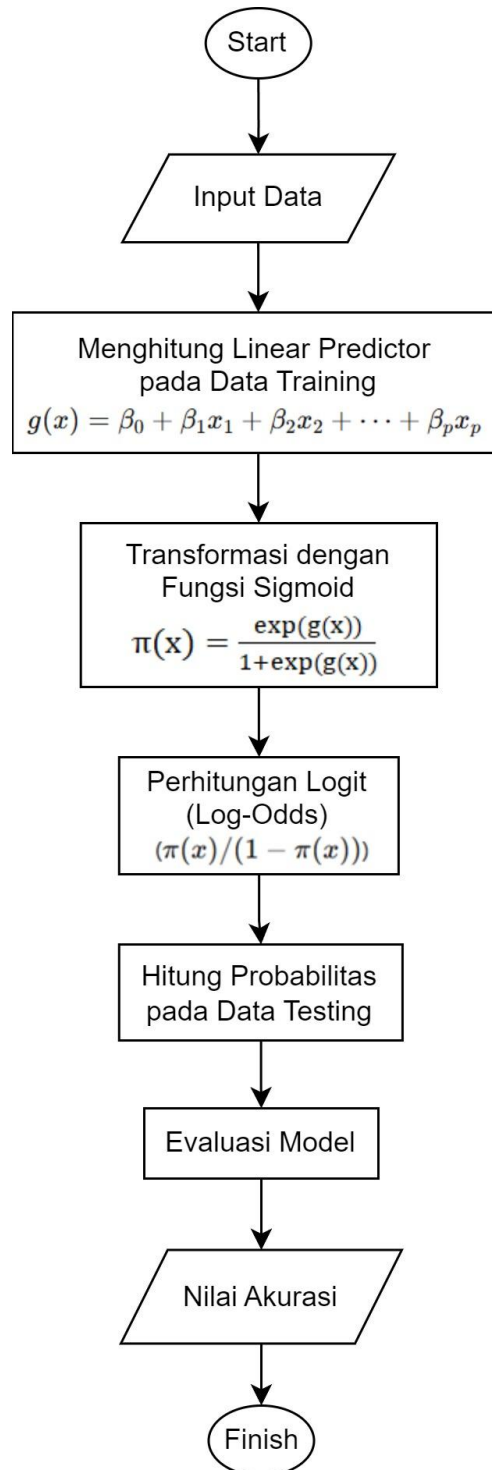


Gambar 5. 1 Arsitektur model *Logistic Regression* pada Klasifikasi Kualitas Perguruan Tinggi

Proses training dari model *Logistic Regression* (LR) dimulai dengan memanggil data *training* yang berisi indikator-indikator kualitas perguruan tinggi yang telah disiapkan sebelumnya. Pada tahap awal, dilakukan penentuan nilai hyperparameter, seperti *learning rate* serta pengaturan regularisasi L1 dan L2, kemudian dilanjutkan dengan inisialisasi nilai koefisien dan *intercept* yang akan diperbarui selama proses *training* berlangsung.

Selanjutnya, model *Logistic Regression* membangun fungsi logit untuk menghitung probabilitas setiap kategori akreditasi berdasarkan indikator yang diberikan. Proses optimisasi dilakukan dengan meminimalkan nilai *loss function* menggunakan pendekatan *log loss* (*cross-entropy loss*). *Training* akan terus berjalan hingga model mencapai titik konvergensi, yaitu kondisi di mana perubahan nilai koefisien sudah sangat kecil atau kriteria berhenti terpenuhi. Setelah proses *training* selesai, koefisien optimal yang diperoleh dari model *Logistic Regression* disimpan untuk kemudian digunakan dalam proses pengujian

(testing). *Flowchart* pada model *Logistic Regression* untuk menganalisis kualitas perguruan tinggi disajikan pada Gambar 5.2.



Gambar 5.1 *Flowchart Model Logistic Regression*

5.2. Implementasi Model *Logistic Regression*

Dalam penelitian ini, algoritma *Logistic Regression* digunakan untuk membangun model klasifikasi dalam menilai kualitas perguruan tinggi di Indonesia. Pemilihan *Logistic Regression* didasarkan pada kemampuannya yang terbukti efektif dalam menyelesaikan masalah klasifikasi dengan interpretasi yang sederhana namun kuat, terutama pada data dengan struktur tabular dan skala yang tidak terlalu besar (Musa, 2013). *Logistic Regression* juga memiliki keunggulan dalam menghindari overfitting dan mampu memberikan probabilitas prediksi yang jelas, yang sangat berguna untuk analisis keputusan.

Proses implementasi model dimulai dari pengolahan data di Kaggle Editor dengan menggunakan bahasa pemrograman Python. Sebelum model dijalankan, data dibagi menjadi dua bagian, yakni data pelatihan dan data pengujian. Langkah ini bertujuan agar model dapat belajar dari data pelatihan, lalu dievaluasi performanya menggunakan data pengujian yang belum pernah dilihat sebelumnya. Tujuan utama pembagian ini adalah untuk memastikan kemampuan model dalam melakukan generalization, yaitu menerapkan pola yang telah dipelajari terhadap data baru secara akurat.

Setelah pembagian data dilakukan, model *Logistic Regression* diinisialisasi dengan beberapa parameter utama. Parameter `class_weight='balanced'` digunakan untuk mengatasi potensi ketidakseimbangan jumlah data antar kelas. Hal ini penting, karena dalam data akreditasi, kemungkinan terdapat lebih banyak perguruan tinggi yang berada pada kategori "Baik" dibandingkan dengan

"Unggul". Dengan menggunakan pengaturan ini, model secara otomatis menyesuaikan bobot setiap kelas berdasarkan frekuensinya, sehingga tidak bias terhadap kelas mayoritas.

Tabel 5. 1 Parameter model *Logistic Regression*

Parameter	Nilai	Keterangan	Penjelasan
class_weight	'balanced'	Bobot Kelas	Mengatur bobot kelas secara otomatis berdasarkan proporsi jumlah sampel pada tiap kelas, untuk mengatasi ketidakseimbangan kelas.
random_state	42	Pengaturan Angka Acak	Memastikan hasil model tetap konsisten setiap kali pelatihan dilakukan dengan data yang sama.
solver	'liblinear'	Metode Optimasi	Algoritma yang digunakan untuk menyelesaikan fungsi optimasi dalam <i>Logistic Regression</i> ; cocok untuk dataset kecil dan mendukung regularisasi L1 dan L2.
max_iter	50	Iterasi Maksimum	Menentukan jumlah maksimum iterasi yang diperbolehkan selama proses pelatihan agar model mencapai konvergensi.

Parameter `random_state` bernilai 42 disetel untuk memastikan konsistensi hasil setiap kali model dijalankan. Artinya, pelatihan ulang model dengan data yang sama akan menghasilkan hasil yang identik, sehingga proses analisis dapat diandalkan dan direproduksi. Selanjutnya, digunakan `solver='liblinear'`, yakni salah satu algoritma optimasi yang cocok untuk dataset dengan ukuran kecil hingga sedang, serta mendukung regularisasi L1 dan L2. Solver ini sering direkomendasikan ketika peneliti ingin mendapatkan hasil klasifikasi yang stabil dan efisien.

Selain itu, jumlah iterasi maksimum ditetapkan pada `max_iter = 50`. Pengaturan ini memberikan kesempatan bagi model untuk mencapai konvergensi secara optimal, terutama ketika data memiliki tingkat kompleksitas yang tinggi atau memerlukan proses pembelajaran yang lebih mendalam terhadap fitur-fitur prediktor. Dengan demikian, proses pelatihan dapat berjalan hingga model menemukan parameter yang paling stabil tanpa berhenti terlalu dini.

Dalam penelitian ini, data awal dibagi menjadi dua bagian utama: data pelatihan dan data pengujian. Tujuan dari pembagian ini adalah agar model dapat belajar dari sebagian data yang tersedia, lalu diuji kemampuannya menggunakan data yang belum pernah dikenalnya. Pendekatan ini penting dilakukan untuk menilai sejauh mana model mampu bekerja secara efektif ketika dihadapkan pada data baru, sebagaimana yang akan terjadi dalam situasi nyata di lapangan.

Tabel 5. 2 Distribusi pelatihan model *Logistic Regression*

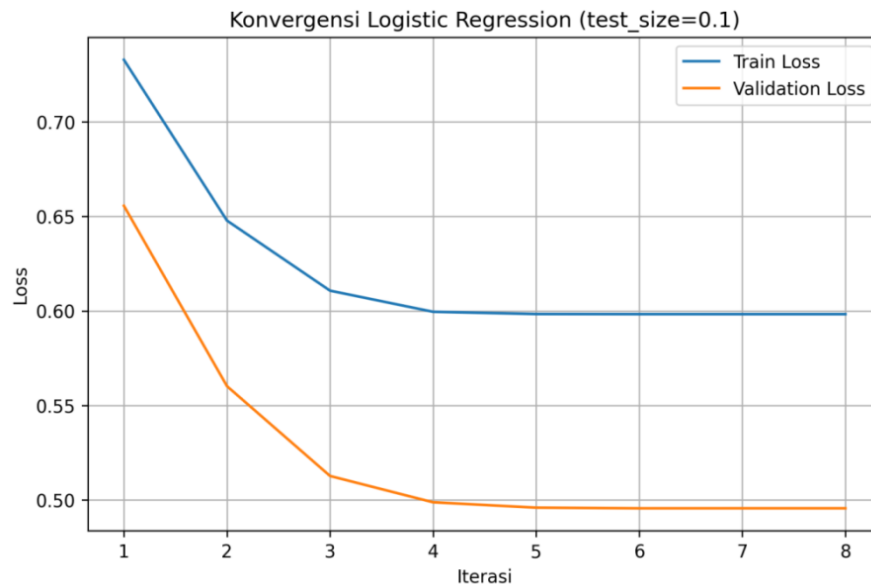
Nama Pelatihan	Rasio Data	Data Latih	Data Uji
LR-A	90:10	106	12
LR-B	80:20	94	24
LR-C	70:30	82	36
LR-D	60:40	70	48
LR-E	50:50	58	60

Untuk memahami bagaimana variasi pembagian data memengaruhi proses pelatihan model pada Tabel 5.2, digunakan beberapa skenario dengan proporsi data pelatihan dan pengujian yang berbeda. Setiap skenario ini diberi kode LR-A, LR-B, LR-C, LR-D, dan LR-E, yang masing-masing merepresentasikan perbedaan dalam komposisi data latih dan data uji. Melalui skenario tersebut, model dilatih untuk memprediksi kualitas perguruan tinggi berdasarkan status Akreditasi, yang menjadi variabel target dalam penelitian ini. Kategori akreditasi yang digunakan meliputi tiga tingkat, yaitu Baik, Baik Sekali, dan Unggul.

5.2.1. Pelatihan Model LR-A

Pada proses eksperimen dari pelatihan model LR-A dalam klasifikasi kualitas perguruan tinggi di Indonesia dengan proporsi data latih sebesar 90% atau sekitar 106 data dan data uji 10% atau sekitar 12 data. Proses pelatihan pada proses eksperimen ini, berhenti dan mencapai titik konvergensi pada iterasi ke-8

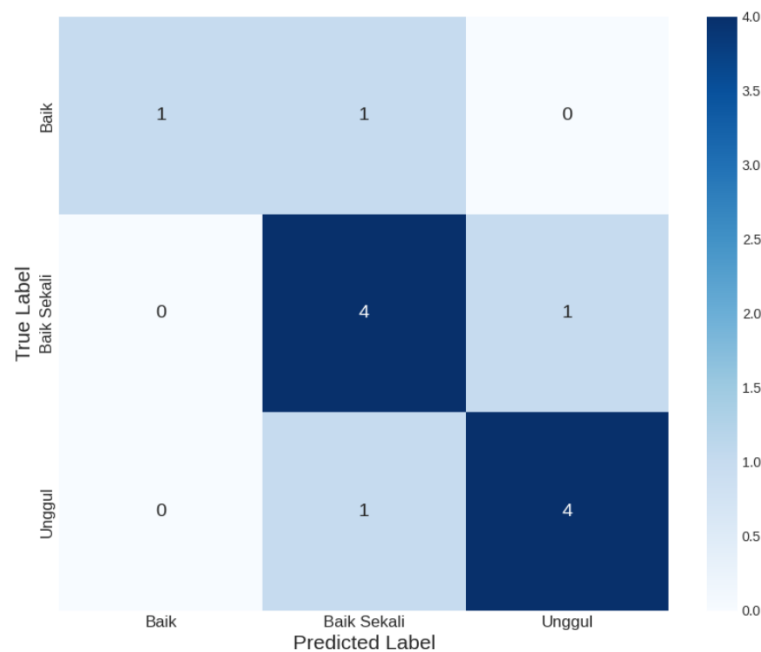
dengan nilai *train loss* sekitar 0.5982 dan nilai *validation loss* sekitar 0.4956 yang disajikan pada Gambar 5.3.



Gambar 5.2 Grafik train dan validation loss model LR-A

Pada Gambar 5.3 menunjukkan nilai *train* dan *validation loss* pada proses pelatihan model LR-A yang terus mengalami penurunan hingga mencapai titik konvergensi pada iterasi ke-8, sehingga model tersebut dianggap telah optimal. Tahap selanjutnya, dilakukan pengujian performa model terhadap data uji. Pengujian performa model pada proses eksperimen ini memperoleh nilai *accuracy* sekitar 58.33%, *precision* sekitar 38.89%, *recall* sekitar 46.67%, dan *F1-Score* sekitar 42.42%.

Selanjutnya, model optimal yang telah didapatkan, kemudian digunakan untuk mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi dengan tiga (3) klasifikasi yaitu “Baik”, “Baik Sekali”, dan “Unggul”. Hasil dari pengujian model LR-A dalam melakukan klasifikasi tingkat akreditasi perguruan tinggi disajikan pada Gambar 5.4.



Gambar 5.3 Hasil klasifikasi model LR-A terhadap tingkat akreditasi

Hasil klasifikasi model LR-A terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 5.4 menampilkan hasil klasifikasi model LR-A terhadap 12 data uji dalam tiga kategori kualitas perguruan tinggi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model hanya mampu mengklasifikasikan 1 dari 2 data dengan benar, sedangkan 1 data lainnya salah diprediksi sebagai Baik Sekali. Pada kategori Baik Sekali, performa model terlihat cukup baik dengan 4 dari 5 data terklasifikasi tepat, meskipun terdapat 1 data yang salah diprediksi sebagai Unggul. Adapun pada kategori Unggul, model berhasil mengenali 4 dari 5 data dengan benar, sementara 1 data lainnya salah diprediksi ke kategori Baik Sekali.

Singkatnya, hasil ini menunjukkan bahwa model LR-A cukup andal dalam mengenali kategori Baik Sekali dan Unggul, namun masih menghadapi tantangan dalam membedakan kategori Baik dengan kelas lain yang berdekatan.

Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-Score* maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 5.3.

Tabel 5. 3 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-A

Kategori	TP	TN	FP	FN
Baik	0	10	0	2
Sangat Baik	3	4	3	2
Unggul	4	5	2	1

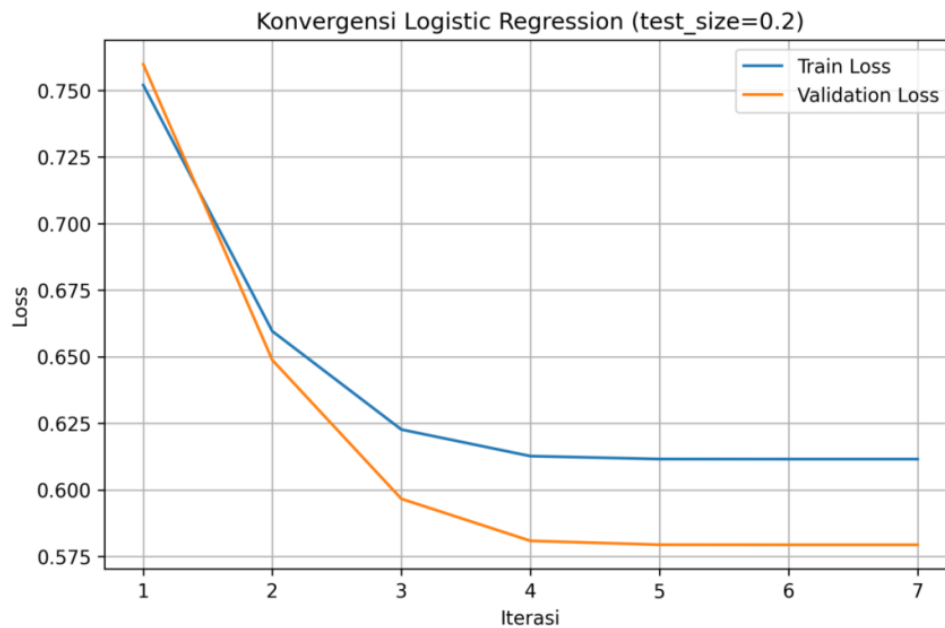
Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang digunakan untuk mengetahui performa model LR-A dalam mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 5.4.

Tabel 5. 4 Performa hasil pengujian model LR-A

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Baik	0.75	1	0.5	0.67	2
Baik Sekali		0.67	0.8	0.73	5
Unggul		0.8	0.8	0.8	5

5.2.2. Pelatihan Model LR-B

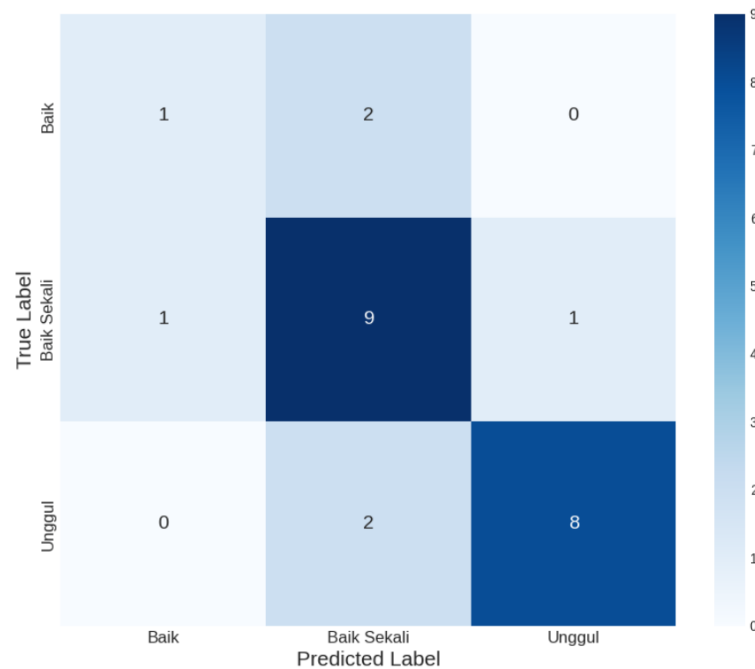
Pada proses eksperimen dari pelatihan model LR-B dalam klasifikasi kualitas perguruan tinggi di Indonesia dengan proporsi data latih sebesar 80% atau sekitar 94 data dan data uji 20% atau sekitar 24 data. Proses pelatihan pada proses eksperimen ini, berhenti dan mencapai titik konvergensi pada iterasi ke-7 dengan nilai *train loss* sekitar 0.6115 dan nilai *validation loss* sekitar 0.5793 yang disajikan pada Gambar 5.5.



Gambar 5.4 Grafik train dan validation loss model LR-B

Pada Gambar 5.5 menunjukkan nilai *train* dan *validation loss* pada proses pelatihan model LR-B yang terus mengalami penurunan hingga mencapai titik konvergensi pada iterasi ke-7, sehingga model tersebut dianggap telah optimal. Tahap selanjutnya, dilakukan pengujian performa model terhadap data uji. Pengujian performa model pada proses eksperimen ini memperoleh nilai *accuracy* sekitar 75%, *precision* sekitar 49.88%, *recall* sekitar 57.58%, dan *F1-Score* sekitar 53.23%.

Selanjutnya, model optimal yang telah didapatkan, kemudian digunakan untuk mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi dengan tiga (3) klasifikasi yaitu “Baik”, “Baik Sekali”, dan “Unggul”. Hasil dari pengujian model LR-B dalam melakukan klasifikasi tingkat akreditasi perguruan tinggi disajikan pada Gambar 5.6.



Gambar 5.5 Hasil klasifikasi model LR-B terhadap tingkat akreditasi

Hasil klasifikasi model LR-B terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 5.6 menyajikan hasil klasifikasi model LR-B terhadap 24 data uji dalam tiga kategori kualitas perguruan tinggi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model hanya mampu mengklasifikasikan 1 dari 3 data dengan benar, sementara 2 data lainnya salah diprediksi ke kategori Baik Sekali. Pada kategori Baik Sekali, performa model terlihat cukup stabil dengan 9 dari 11 data terklasifikasi tepat, sedangkan 1 data salah diprediksi ke kategori Baik dan 1 data lainnya ke kategori Unggul. Adapun pada kategori Unggul, model berhasil mengenali 8 dari 10 data dengan benar, sementara 2 data salah diprediksi sebagai Baik Sekali.

Secara keseluruhan, hasil ini menunjukkan bahwa model LR-B cukup baik dalam mengenali kategori Baik Sekali dan Unggul, meskipun masih mengalami kesulitan dalam membedakan kategori Baik dengan kelas menengah serta

membedakan Unggul dengan kelas di bawahnya. Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-Score* maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 5.5.

Tabel 5. 5 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-B

Kategori	TP	TN	FP	FN
Baik	0	21	0	3
Sangat Baik	8	10	3	3
Unggul	10	11	3	0

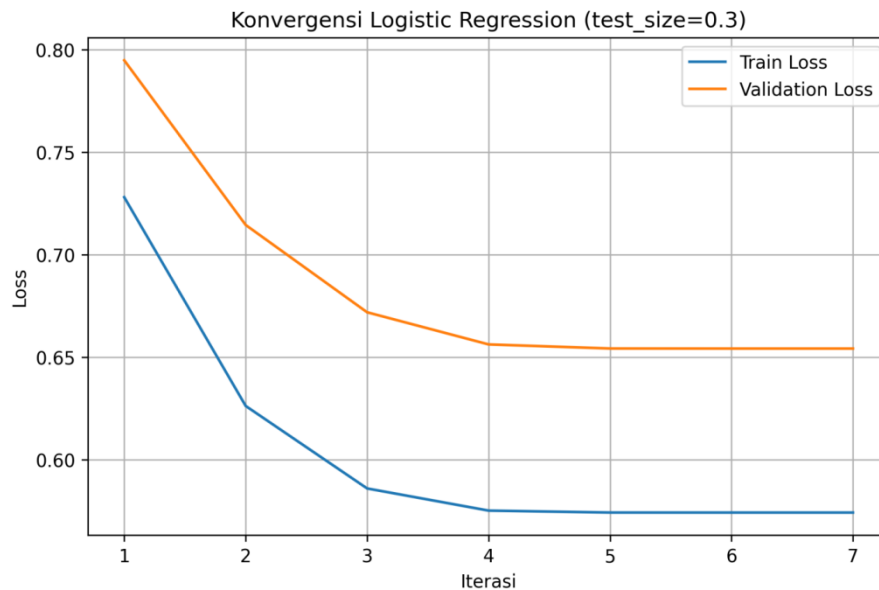
Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang digunakan untuk mengetahui performa model LR-B dalam mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 5.6.

Tabel 5. 6 Performa hasil pengujian model LR-B

Kategori	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Baik	0.75	0.5	0.33	0.4	3
Baik Sekali		0.69	0.82	0.75	11
Unggul		0.89	0.8	0.84	10

5.2.3. Pelatihan Model LR-C

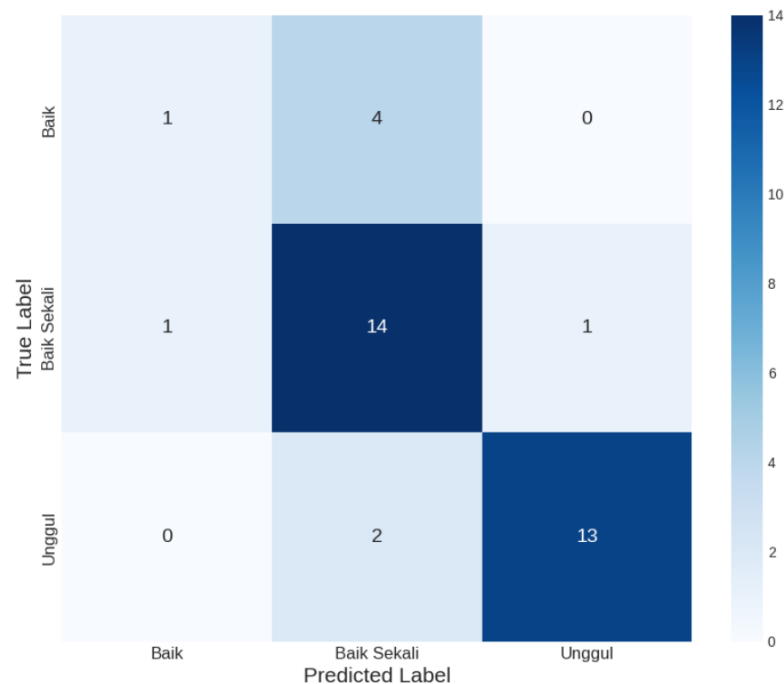
Pada proses eksperimen dari pelatihan model LR-C dalam klasifikasi kualitas perguruan tinggi di Indonesia dengan proporsi data latih sebesar 70% atau sekitar 82 data dan data uji 30% atau sekitar 36 data. Proses pelatihan pada proses eksperimen ini, berhenti dan mencapai titik konvergensi pada iterasi ke-7 dengan nilai *train loss* sekitar 0.5741 dan nilai *validation loss* sekitar 0.6541 yang disajikan pada Gambar 5.7.



Gambar 5.6 Grafik train dan validation loss model LR-C

Pada Gambar 5.7 menunjukkan nilai *train* dan *validation loss* pada proses pelatihan model LR-C yang terus mengalami penurunan hingga mencapai titik konvergensi pada iterasi ke-7, sehingga model tersebut dianggap telah optimal. Tahap selanjutnya, dilakukan pengujian performa model terhadap data uji. Pengujian performa model pada proses eksperimen ini memperoleh nilai *accuracy* sekitar 72.22%, *precision* sekitar 49.78%, *recall* sekitar 55.83%, dan *F1-Score* sekitar 52.15%.

Selanjutnya, model optimal yang telah didapatkan, kemudian digunakan untuk mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi dengan tiga (3) klasifikasi yaitu “Baik”, “Baik Sekali”, dan “Unggul”. Hasil dari pengujian model LR-C dalam melakukan klasifikasi tingkat akreditasi perguruan tinggi disajikan pada Gambar 5.8.



Gambar 5.7 Hasil klasifikasi model LR-C terhadap tingkat akreditasi

Hasil klasifikasi model LR-C terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 5.8 menyajikan hasil klasifikasi model LR-C terhadap 36 data uji dalam tiga kategori mutu perguruan tinggi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model hanya mampu mengklasifikasikan 1 dari 5 data dengan benar, sementara 4 data lainnya salah diprediksi sebagai Baik Sekali. Pada kategori Baik Sekali, performa model terlihat sangat baik dengan 14 dari 16 data berhasil diklasifikasikan secara tepat, meskipun masih terdapat 1 data yang salah ke kategori Baik dan 1 data lainnya ke kategori Unggul. Adapun pada kategori Unggul, model mampu mengenali 13 dari 15 data dengan benar, sedangkan 2 data lainnya salah diprediksi sebagai Baik Sekali.

Secara keseluruhan, hasil ini menunjukkan bahwa model LR-C cukup andal dalam membedakan kategori Baik Sekali dan Unggul, namun masih menghadapi

kesulitan dalam mengenali kategori Baik secara konsisten karena adanya kemiripan ciri antar kelas. Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-Score* maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 5.7.

Tabel 5. 7 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-C

Kategori	TP	TN	FP	FN
Baik	1	30	1	4
Baik Sekali	14	14	6	2
Unggul	13	20	1	2

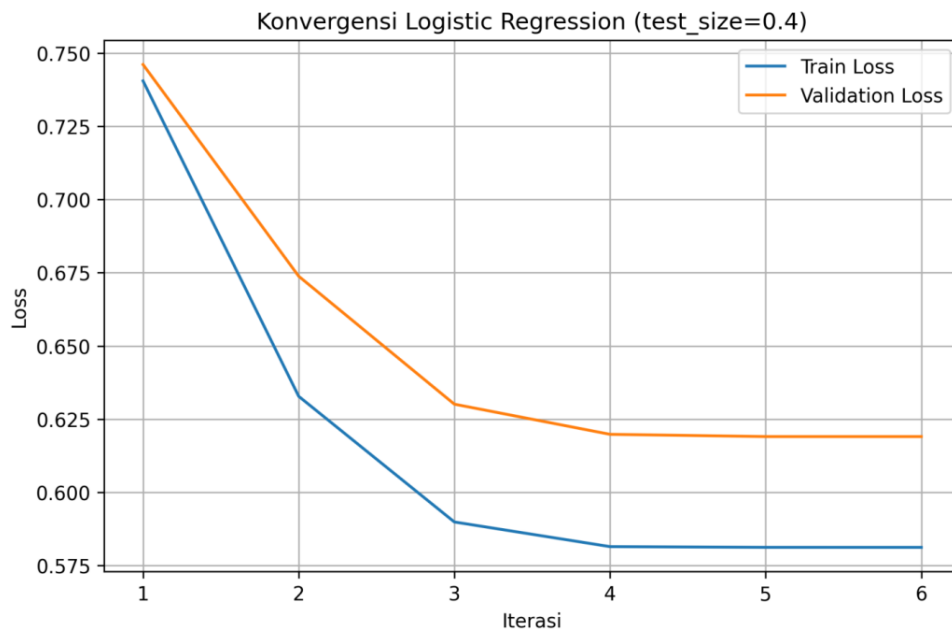
Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang digunakan untuk mengetahui performa model LR-C dalam mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 5.8.

Tabel 5. 8 Performa hasil pengujian model LR-C

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Baik	0.78	0.5	0.2	0.29	5
Baik Sekali		0.7	0.88	0.78	16
Unggul		0.93	0.87	0.9	15

5.2.4. Pelatihan Model LR-D

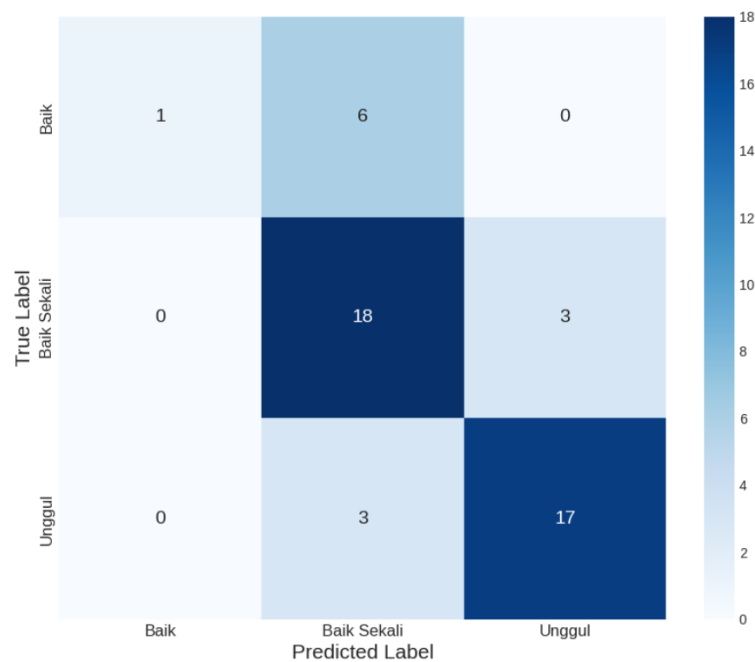
Pada proses eksperimen dari pelatihan model LR-D dalam klasifikasi kualitas perguruan tinggi di Indonesia dengan proporsi data latih sebesar 60% atau sekitar 70 data dan data uji 40% atau sekitar 48 data. Proses pelatihan pada proses eksperimen ini, berhenti dan mencapai titik konvergensi pada iterasi ke-6 dengan nilai *train loss* sekitar 0.5812 dan nilai *validation loss* sekitar 0.6190 yang disajikan pada Gambar 5.9.



Gambar 5.8 Grafik train dan validation loss model LR-D

Pada Gambar 5.9 menunjukkan nilai *train* dan *validation loss* pada proses pelatihan model LR-D yang terus mengalami penurunan hingga mencapai titik konvergensi pada iterasi ke-6, sehingga model tersebut dianggap telah optimal. Tahap selanjutnya, dilakukan pengujian performa model terhadap data uji. Pengujian performa model pada proses eksperimen ini memperoleh nilai *accuracy* sekitar 72.92%, *precision* sekitar 49.91%, *recall* sekitar 55.45%, dan *F1-Score* sekitar 52.19%.

Selanjutnya, model optimal yang telah didapatkan, kemudian digunakan untuk mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi dengan tiga (3) klasifikasi yaitu “Baik”, “Baik Sekali”, dan “Unggul”. Hasil dari pengujian model LR-D dalam melakukan klasifikasi tingkat akreditasi perguruan tinggi disajikan pada Gambar 5.10.



Gambar 5.9 Hasil klasifikasi model LR-D terhadap tingkat akreditasi

Hasil klasifikasi model LR-D terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 5.10 menunjukkan hasil klasifikasi model LR-D terhadap 48 data uji dalam tiga kategori mutu perguruan tinggi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model hanya mampu mengklasifikasikan 1 dari 7 data dengan benar, sementara 6 data lainnya salah diprediksi ke kategori Baik Sekali. Pada kategori Baik Sekali, performa model cukup dominan dengan 18 dari 21 data berhasil diklasifikasikan secara tepat, meskipun 3 data lainnya salah diprediksi ke kategori Unggul. Adapun pada kategori Unggul, model mampu mengenali 17 dari 20 data dengan benar, sedangkan 3 data lainnya salah diprediksi sebagai Baik Sekali.

Secara keseluruhan, hasil tersebut menunjukkan bahwa model LR-D cukup andal dalam mengenali kategori Baik Sekali dan Unggul, namun masih

menghadapi tantangan dalam membedakan kategori Baik karena kecenderungan misklasifikasi ke kelas menengah. Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-Score* maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 5.9.

Tabel 5. 9 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-D

Kategori	TP	TN	FP	FN
Baik	0	42	0	6
Sangat Baik	20	16	10	2
Unggul	16	26	2	4

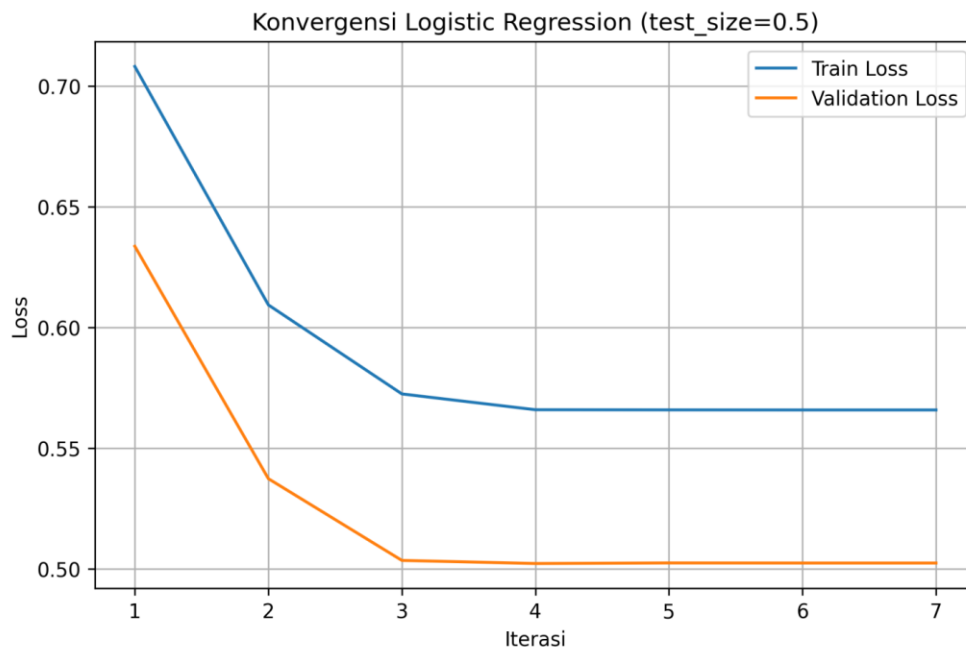
Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang digunakan untuk mengetahui performa model LR-D dalam mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 5.10.

Tabel 5. 10 Performa hasil pengujian model LR-D

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Baik	0.75	1	0.14	0.25	7
Baik Sekali		0.67	0.86	0.75	21
Unggul		0.85	0.85	0.85	20

5.2.5. Pelatihan Model LR-E

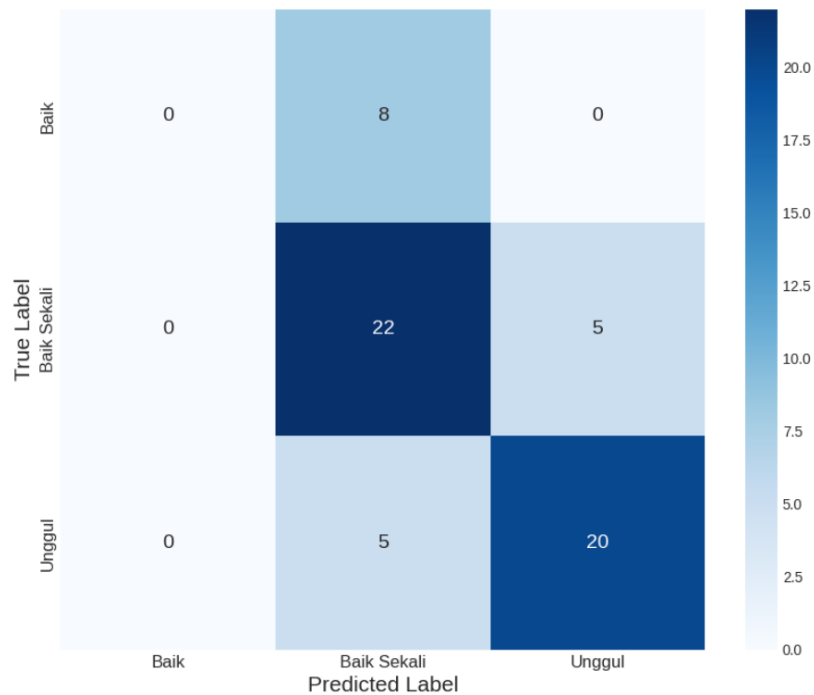
Pada proses eksperimen dari pelatihan model LR-E dalam klasifikasi kualitas perguruan tinggi di Indonesia dengan proporsi data latih sebesar 50% atau sekitar 58 data dan data uji 50% atau sekitar 60 data. Proses pelatihan pada proses eksperimen ini, berhenti dan mencapai titik konvergensi pada iterasi ke-7 dengan nilai *train loss* sekitar 0.5657 dan nilai *validation loss* sekitar 0.5023 yang disajikan pada Gambar 5.11.



Gambar 5.10 Grafik train dan validation loss model LR-E

Pada Gambar 5.11 menunjukkan nilai *train* dan *validation loss* pada proses pelatihan model LR-E yang terus mengalami penurunan hingga mencapai titik konvergensi pada iterasi ke-7, sehingga model tersebut dianggap telah optimal. Tahap selanjutnya, dilakukan pengujian performa model terhadap data uji. Pengujian performa model pada proses eksperimen ini memperoleh nilai *accuracy* sekitar 69.49%, *precision* sekitar 47.25%, *recall* sekitar 53.59%, dan *F1-Score* sekitar 50%.

Selanjutnya, model optimal yang telah didapatkan, kemudian digunakan untuk mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi dengan tiga (3) klasifikasi yaitu “Baik”, “Baik Sekali”, dan “Unggul”. Hasil dari pengujian model LR-E dalam melakukan klasifikasi tingkat akreditasi perguruan tinggi disajikan pada Gambar 5.12.



Gambar 5.11 Hasil klasifikasi model LR-E terhadap tingkat akreditasi

Hasil klasifikasi model LR-E terhadap tingkat akreditasi perguruan tinggi di Indonesia memiliki nilai metrik yang berbeda-beda. Pada Gambar 5.12 menunjukkan hasil klasifikasi model LR-E terhadap 60 data uji dalam tiga kategori kualitas perguruan tinggi, yaitu Baik, Baik Sekali, dan Unggul. Pada kategori Baik, model gagal sepenuhnya mengenali kelas ini, di mana seluruh 8 data aktual salah diprediksi sebagai Baik Sekali. Pada kategori Baik Sekali, performa model terlihat cukup dominan dengan 22 dari 27 data berhasil diklasifikasikan dengan benar, sementara 5 data lainnya salah diprediksi sebagai Unggul. Adapun pada kategori Unggul, model mampu mengklasifikasikan 20 dari 25 data dengan tepat, sedangkan 5 data lainnya salah diprediksi ke kategori Baik Sekali.

Secara keseluruhan, hasil tersebut menunjukkan bahwa model LR-E andal dalam mengenali kategori Baik Sekali dan Unggul, tetapi sepenuhnya gagal dalam

mengklasifikasikan kategori Baik, yang menandakan adanya tantangan serius terkait ketidakseimbangan kelas serta kemiripan fitur antar kategori. Selanjutnya, untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-Score* maka dibutuhkan nilai TP, TN, FP, dan FN dari hasil klasifikasi tingkat kualitas perguruan tinggi berdasarkan tingkat akreditasi yang secara rinci disajikan dalam Tabel 5.11.

Tabel 5. 11 Nilai TP, TN, FP, dan FN setiap kategori pada model LR-E

Kategori	TP	TN	FP	FN
Baik	0	51	0	8
Sangat Baik	20	21	12	6
Unggul	21	28	6	4

Setelah didapatkan nilai TP, TN, FP, dan FN dari setiap kategori akreditasi perguruan tinggi, langkah selanjutnya adalah menghitung metrik evaluasi yang digunakan untuk mengetahui performa model LR-E dalam mengukur kualitas perguruan tinggi di Indonesia berdasarkan tingkat akreditasi yang secara rinci disajikan pada Tabel 5.12.

Tabel 5. 12 Performa hasil pengujian model LR-E

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Baik	0.75	0	0	0	8
Baik Sekali		0.63	0.81	0.71	27
Unggul		0.80	0.80	0.80	25

5.3. Ujicoba Model *Logistic Regression*

Setelah proses pelatihan selesai, masing-masing model *Logistic Regression* diujicoba menggunakan data uji untuk menilai kinerjanya dalam mengklasifikasikan kualitas perguruan tinggi ke dalam tiga kategori: Baik, Baik Sekali, dan Unggul. Evaluasi dilakukan dengan menggunakan sejumlah metrik penting, seperti *accuracy*, *precision*, *recall*, dan *F1-Score*, yang masing-masing memberikan sudut pandang berbeda terhadap performa model.

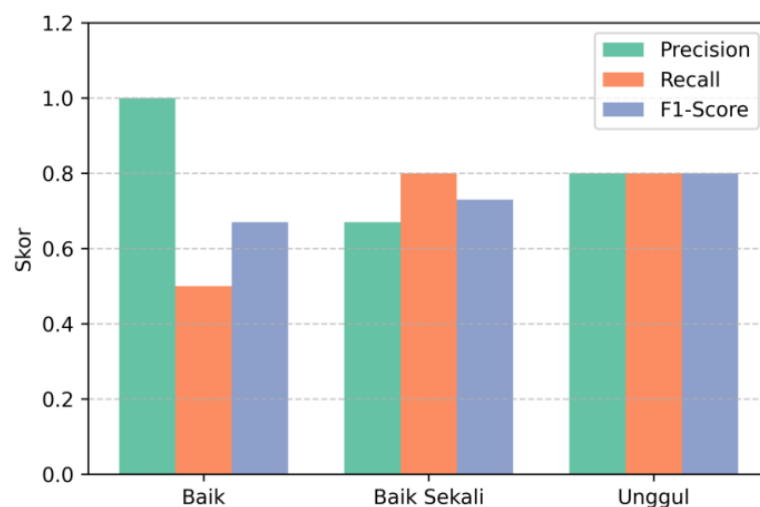
Tabel 5. 13 Hasil ujicoba klasifikasi kualitas model *Logistic Regression*

Model	Kategori	Accuracy	Precision	Recall	F1-Score	Support
LR-A	Baik	0.75	1	0.5	0.67	2
	Baik Sekali		0.67	0.8	0.73	5
	Unggul		0.8	0.8	0.8	5
LR-B	Baik	0.75	0.5	0.33	0.4	3
	Baik Sekali		0.69	0.82	0.75	11
	Unggul		0.89	0.8	0.84	10
LR-C	Baik	0.78	0.5	0.2	0.29	5
	Baik Sekali		0.7	0.88	0.78	16
	Unggul		0.93	0.87	0.9	15
LR-D	Baik	0.75	1	0.14	0.25	7
	Baik Sekali		0.67	0.86	0.75	21
	Unggul		0.85	0.85	0.85	20
LR-E	Baik	0.75	0	0	0	8
	Baik Sekali		0.63	0.81	0.71	27
	Unggul		0.80	0.80	0.80	25

Pada Tabel 5.13 menunjukkan bahwa model LR-A menghasilkan nilai accuracy sebesar 0.75, yang berarti sekitar 75% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Untuk kategori Baik mencatat precision sempurna 1.00 atau sekitar 100%, yang berarti seluruh prediksi benar-benar tepat. Namun, recall hanya 0.50 atau sekitar 50%, sehingga separuh data aktual kategori ini tidak terdeteksi. Dengan f1-score 0.67 atau sekitar 67%, performa model masih terbatas, terlebih karena support hanya 2 data sehingga hasil kurang stabil. Kategori Baik Sekali menunjukkan kinerja lebih seimbang dengan precision 0.67 atau sekitar 67%, recall 0.80 atau sekitar 80%, dan f1-score 0.73 atau sekitar 73%, didukung oleh 5 data uji yang meski terbatas, cukup representatif. Sementara itu, kategori Unggul memperlihatkan performa paling konsisten dengan precision 0.80 atau sekitar 80%, recall 0.80 atau sekitar 80%, dan f1-score 0.80 atau sekitar 80%, yang menandakan model LR-A efektif dan

andal dalam mengidentifikasi kategori tertinggi ini, terutama dengan dukungan 5 data yang memberikan dasar penilaian memadai.

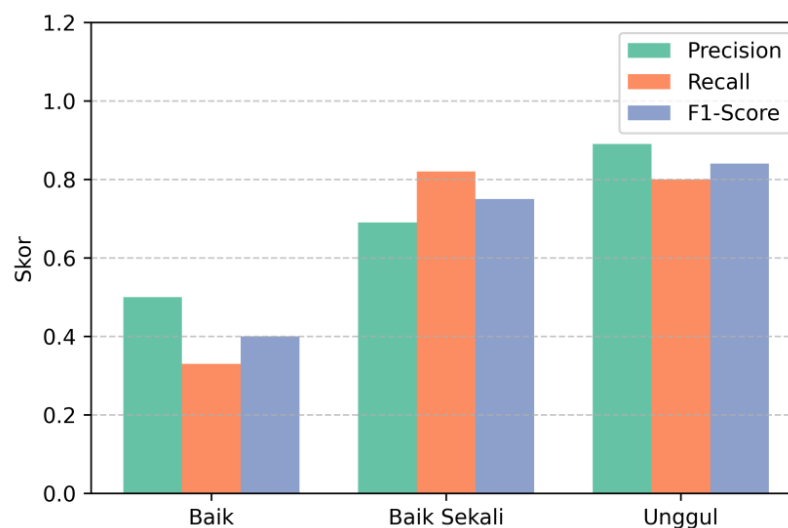
Secara keseluruhan, evaluasi terhadap model LR-A pada Gambar 5.12 menunjukkan performa klasifikasi yang baik dan stabil, terutama dalam konteks data uji yang terbatas. Model ini layak dijadikan alternatif dalam proses klasifikasi, terutama ketika diperlukan pendekatan yang efisien namun tetap menjaga keseimbangan antar kelas. Untuk mencapai performa yang lebih optimal, disarankan untuk menerapkan teknik seperti tuning parameter, penyeimbangan kelas, atau penambahan fitur yang lebih informatif terhadap pola klasifikasi.



Gambar 5.12 Perbandingan matrik evaluasi model LR-A

Selanjutnya, pada Tabel 5.13 menunjukkan bahwa model LR-B menghasilkan nilai accuracy sebesar 0.75, yang berarti sekitar 75% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Untuk kategori Baik menunjukkan performa rendah dengan precision 0.50 atau sekitar 50%, recall 0.33 atau sekitar 33%, dan f1-score 0.40 atau sekitar 40%. Hasil ini menandakan bahwa model kesulitan mengenali kelas ini, terlebih karena jumlah

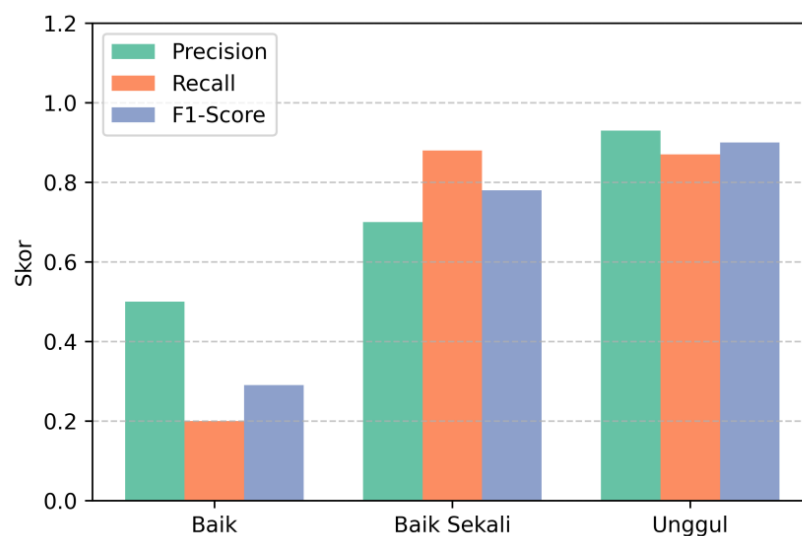
data uji sangat terbatas (3 data) sehingga evaluasinya sensitif terhadap kesalahan. Kategori Baik Sekali tampil lebih baik dengan precision 0.69 atau sekitar 69%, recall 0.82 atau sekitar 82%, dan f1-score 0.75 atau sekitar 75%, menunjukkan keseimbangan yang cukup andal, terutama dengan dukungan 11 data uji yang lebih memadai. Sementara itu, kategori Unggul memperlihatkan performa terbaik, dengan precision 0.89 atau sekitar 89%, recall 0.80 atau sekitar 80%, dan f1-score 0.84 atau sekitar 84%, yang mencerminkan efektivitas dan konsistensi model LR-B dalam mengklasifikasikan kelas tertinggi ini, didukung oleh 10 data uji.



Gambar 5.13 Perbandingan matrik evaluasi model LR-B

Secara keseluruhan, hasil evaluasi terhadap model LR-B pada Gambar 5.13 menunjukkan bahwa model ini layak dipertimbangkan sebagai salah satu kandidat untuk tugas klasifikasi, terutama ketika dibutuhkan efisiensi dan hasil yang cukup merata. Meski demikian, diperlukan langkah-langkah peningkatan, seperti penyeimbangan distribusi kelas, tuning hyperparameter, serta eksplorasi fitur yang lebih relevan untuk mengoptimalkan performa model dalam konteks klasifikasi yang lebih kompleks.

Selanjutnya, pada Tabel 5.13 menunjukkan bahwa model LR-C menghasilkan nilai akurasi sebesar 0.78, yang berarti sekitar 77.78% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Untuk kategori Baik, model mencatat precision sebesar 0.50 atau sekitar 50%, recall 0.20 atau sekitar 20%, dan f1-score 0.29 atau sekitar 29%. Hasil ini menunjukkan performa rendah dengan dukungan data sebanyak 5. Di sisi lain, kategori Baik Sekali lebih baik dengan precision 0.70 atau sekitar 70%, recall 0.88 atau sekitar 88%, serta f1-score 0.78 atau sekitar 78%, didukung oleh 16 data uji yang membuat hasilnya lebih representatif. Sementara itu, kategori Unggul menjadi yang terbaik, ditandai precision 0.93 atau sekitar 93%, recall 0.87 atau sekitar 87%, dan f1-score 0.90 atau sekitar 90%, dengan support sebanyak 15 data, sehingga menegaskan konsistensi model LR-C dalam mengenali mutu tertinggi.



Gambar 5.14 Perbandingan matrik evaluasi model LR-C

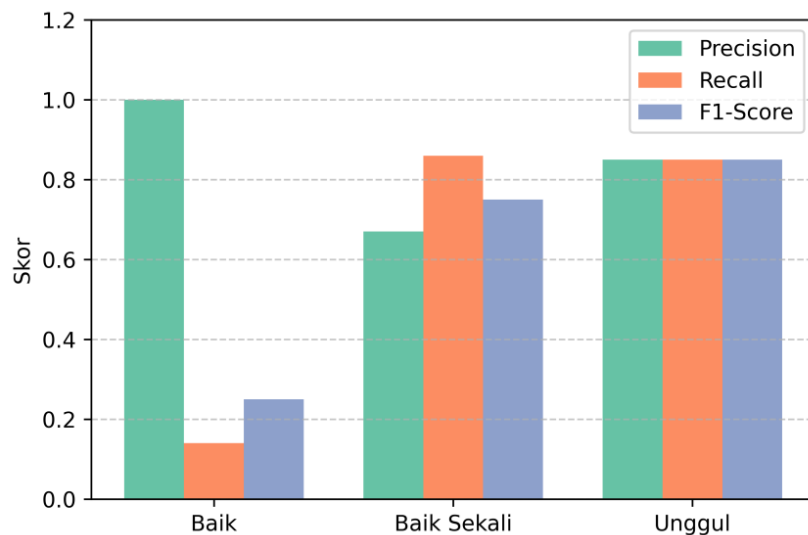
Secara keseluruhan, evaluasi terhadap model LR-C pada Gambar 5.14 menunjukkan bahwa model ini memiliki performa yang cukup baik dan dapat diandalkan untuk tugas klasifikasi. Nilai akurasi yang relatif tinggi serta stabilitas

pada metrik weighted average menjadi indikator bahwa model ini layak digunakan, khususnya dalam konteks data dengan distribusi kelas yang tidak terlalu ekstrem. Untuk lebih mengoptimalkan kinerjanya, disarankan untuk melakukan penyesuaian parameter, meningkatkan kualitas fitur, atau mengimplementasikan metode penyeimbangan kelas jika diperlukan.

Tabel 5.13 menunjukkan bahwa model LR-D menghasilkan nilai akurasi sebesar 0.75, yang berarti sekitar 75% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Kategori Baik, model mencatat precision sempurna sebesar 1.00 atau sekitar 100%, namun recall hanya 0.14 atau sekitar 14%. Nilai f1-score sebesar 0.25 atau sekitar 25% menunjukkan performa masih rendah, dengan jumlah support sebanyak 7 data. Sementara itu, kategori Baik Sekali menunjukkan performa yang lebih stabil. Dengan precision sebesar 0.67 atau sekitar 67% dan recall sebesar 0.86 atau sekitar 86%, nilai f1-score mencapai 0.75 atau sekitar 75%. Dengan support sebanyak 21 data, kinerja model pada kelas ini tergolong cukup andal. Selain itu, untuk kategori Unggul, model memperlihatkan hasil terbaik. Dengan precision sebesar 0.85 atau sekitar 85% dan recall sebesar 0.85 atau sekitar 85%, nilai f1-score juga sebesar 0.85 atau sekitar 85%. Dengan jumlah *support* sebanyak 20 data, performa model dapat dianggap kuat dan konsisten pada kelas ini.

Secara keseluruhan, evaluasi terhadap model LR-D pada Gambar 5.15 menegaskan bahwa model ini memiliki kinerja klasifikasi yang baik dalam konteks umum, terutama dalam hal ketepatan prediksi. Akan tetapi, untuk meningkatkan cakupan dan kesetaraan kinerja di seluruh kelas, disarankan agar

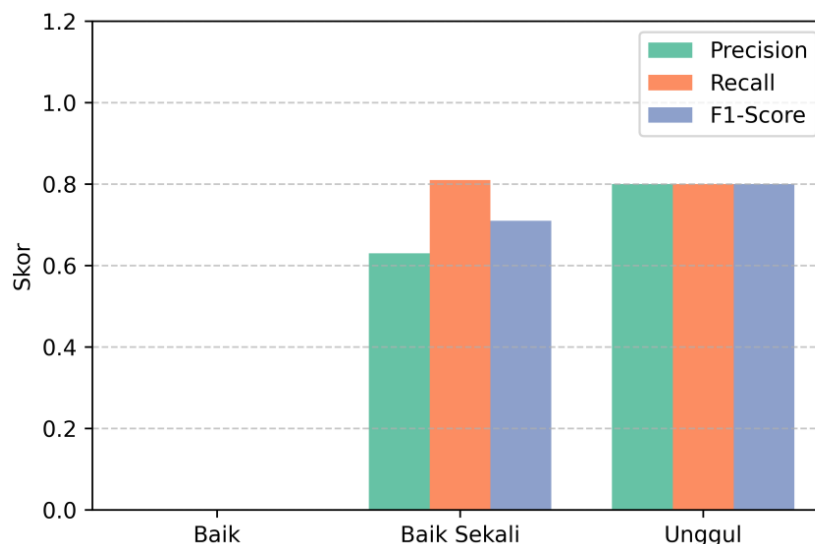
dilakukan strategi peningkatan seperti penyeimbangan kelas, tuning parameter, atau pemilihan fitur yang lebih relevan agar performa model dapat lebih optimal dan adil dalam semua kategori.



Gambar 5.15 Perbandingan matrik evaluasi model LR-D

Demikian, pada Tabel 5.13 menunjukkan bahwa model LR-E menghasilkan nilai accuracy sebesar 0.70, yang berarti sekitar 70% dari total prediksi model sesuai dengan label aktual dari masing-masing kategori. Untuk kategori Baik, model gagal sepenuhnya melakukan klasifikasi. Baik precision, recall, maupun f1-score semuanya bernilai 0.00 atau sekitar 0%, menandakan bahwa tidak ada dari 8 data aktual kelas ini yang berhasil dikenali oleh model. Sementara itu, kategori Baik Sekali menjadi titik prediksi dominan. Dengan precision sebesar 0.63 atau sekitar 63% dan recall sebesar 0.81 atau sekitar 81%, nilai f1-score mencapai 0.71 atau sekitar 71%. Dengan support sebanyak 27 data, performa model pada kelas ini cukup representatif, meskipun terdapat kecenderungan middle bias. Selain itu, untuk kategori Unggul, model menunjukkan kinerja terbaik. Precision dan recall sama-sama sebesar 0.80 atau sekitar 80%, menghasilkan f1-score sebesar 0.80

atau sekitar 80%. Dengan support sebanyak 25 data, hasil ini menunjukkan kemampuan model yang konsisten dalam mengenali kelas tertinggi.



Gambar 5.16 Perbandingan matrik evaluasi model LR-E

Evaluasi terhadap model LR-E pada Gambar 5.16 menunjukkan bahwa model ini memiliki akurasi yang layak, namun belum cukup kuat dalam menjaga keseimbangan performa di seluruh kategori. Untuk meningkatkan hasil klasifikasi, disarankan agar dilakukan beberapa langkah strategis seperti penyeimbangan distribusi kelas, tuning hyperparameter, serta pengayaan fitur yang lebih relevan agar performa model menjadi lebih merata dan optimal.

Berdasarkan hasil ujicoba klasifikasi kualitas perguruan tinggi pada lima skenario eksperimen model *Logistic Regression* (LR-A hingga LR-E), terlihat bahwa variasi rasio data latih dan uji memberikan dampak yang cukup signifikan terhadap performa klasifikasi model. Setiap skenario menunjukkan karakteristik hasil yang berbeda, dengan keunggulan dan keterbatasan masing-masing dalam mengelompokkan mutu perguruan tinggi ke dalam tiga kelas utama: Baik, Baik Sekali, dan Unggul. *Logistic Regression*, sebagai model linear berbasis

probabilistik, menunjukkan sensitivitas terhadap distribusi data, sehingga perubahan proporsi data pelatihan dan pengujian sangat mempengaruhi akurasi dalam proses pengukuran kualitas perguruan tinggi.

Tabel 5. 14 Performa model *Logistic Regression*

Nama Pelatihan	Rasio Data	Nilai Akurasi
LR-A	90:10	75%
LR-B	80:20	75%
LR-C	70:30	77.78%
LR-D	60:40	75%
LR-E	50:50	70%

Pada Tabel 5.14 menunjukkan bahwa model LR-A, yang menggunakan rasio data latih dan uji sebesar 90:10, mencatatkan akurasi sebesar 75%. Meskipun jumlah data latih yang besar dapat membantu model belajar lebih banyak pola, ukuran data uji yang terlalu kecil dapat menyebabkan hasil evaluasi menjadi kurang representatif terhadap performa sebenarnya. Hal ini menunjukkan bahwa meskipun akurasinya terlihat cukup baik, stabilitas prediksi perlu dikaji lebih lanjut, terutama untuk memastikan bahwa model tidak mengalami overfitting.

Model LR-B, dengan rasio data 80:20, mempertahankan nilai akurasi yang sama, yakni 75%. Konsistensi ini menunjukkan bahwa peningkatan jumlah data uji belum memberikan perubahan signifikan terhadap performa klasifikasi, namun memberikan hasil evaluasi yang lebih stabil dan dapat diandalkan untuk menilai generalisasi model.

Selanjutnya, model LR-C yang menggunakan rasio 70:30 menunjukkan performa terbaik di antara seluruh skenario, dengan akurasi tertinggi mencapai 77,78%. Rasio ini memberikan keseimbangan optimal antara proses pelatihan dan pengujian, memungkinkan model *Logistic Regression* menangkap pola klasifikasi

secara lebih menyeluruh tanpa kehilangan akurasi akibat keterbatasan data uji atau pelatihan. Di sisi lain, model LR-D, yang menggunakan rasio data latih dan uji sebesar 60:40, mengalami sedikit penurunan akurasi kembali ke angka 75%. Meskipun jumlah data uji meningkat, yang seharusnya dapat membantu dalam evaluasi yang lebih akurat, penurunan jumlah data latih bisa menyebabkan berkurangnya efektivitas model dalam mengenali pola klasifikasi, terutama untuk kategori mutu yang memiliki jumlah data lebih sedikit.

Terakhir, model LR-E dengan rasio seimbang 50:50 menunjukkan akurasi terendah yaitu sebesar 70%. Penurunan ini mengindikasikan bahwa saat data pelatihan terlalu sedikit, *Logistic Regression* cenderung kehilangan kemampuan dalam membangun model klasifikasi yang kuat. Hal ini juga menunjukkan potensi risiko underfitting, di mana model tidak mampu menangkap kompleksitas hubungan antar variabel secara optimal.

Secara keseluruhan, model **LR-C** memberikan performa terbaik dengan nilai akurasi sekitar 77,78%, diikuti oleh LR-A, LR-B, dan LR-D yang sama-sama mencatat nilai akurasi sekitar 75%. Sementara itu, LR-E menempati posisi terendah dengan nilai akurasi sekitar 70%. Temuan ini menegaskan pentingnya pemilihan rasio data latih dan uji yang seimbang tidak hanya dari sisi kuantitas, tetapi juga kualitas distribusi data dalam membangun model klasifikasi yang andal untuk mengukur kualitas perguruan tinggi.

Tabel 5. 15 Nilai TP, TN, FP, FN model LR-C

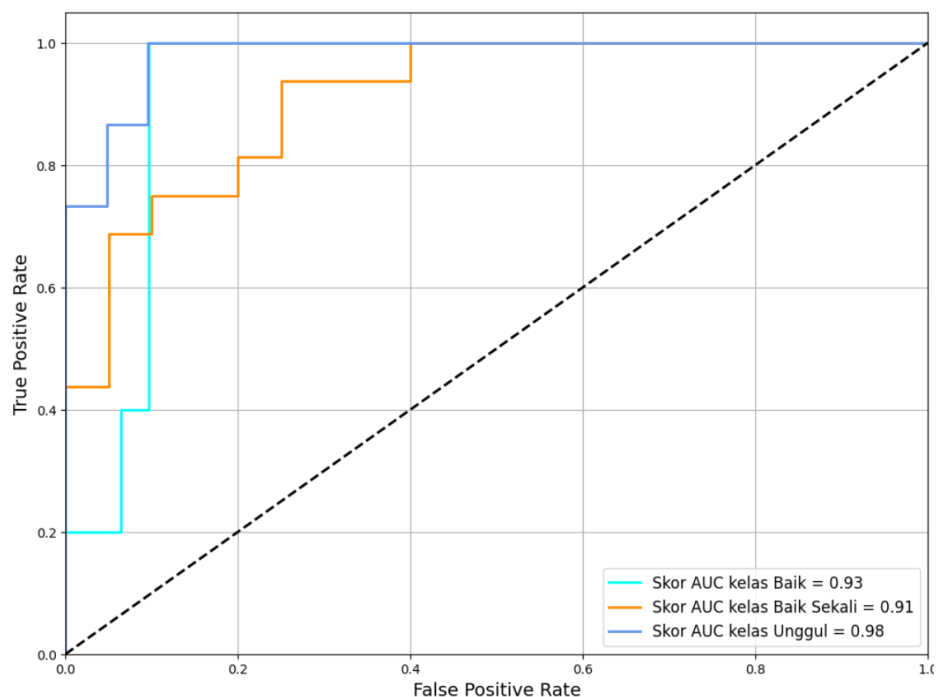
Kelas	True Positive	True Negative	False Positive	False Negative
Baik	1	30	1	4
Baik Sekali	14	14	6	2
Unggul	13	20	1	2

Tabel 5.15 menyajikan rincian performa model LR-C dalam proses pengukuran kualitas perguruan tinggi ke dalam tiga kategori, yakni Baik, Baik Sekali, dan Unggul, berdasarkan metrik True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Analisis ini memberikan pemahaman yang lebih mendalam mengenai sejauh mana model *Logistic Regression* dapat mengenali dan membedakan antar kelas mutu perguruan tinggi.

Pada kategori Baik, model hanya mampu mengidentifikasi 1 data dengan benar (TP), dan 30 data lainnya diklasifikasikan secara tepat sebagai bukan Baik (TN). Namun demikian, terdapat 4 data yang seharusnya masuk ke dalam kategori ini tetapi gagal dikenali (FN), serta 1 data yang salah diklasifikasikan sebagai Baik (FP). Jumlah TP yang sangat rendah dan FN yang tinggi menunjukkan bahwa model mengalami kesulitan dalam mengenali ciri khas dari kategori ini, yang mengakibatkan rendahnya recall dan menunjukkan bahwa model cenderung under-detect terhadap mutu Baik.

Sementara itu, pada kategori Baik Sekali, model berhasil mengklasifikasikan 14 data dengan benar (TP) dan 14 data lainnya dikenali secara tepat sebagai bukan kelas ini (TN). Akan tetapi, masih terdapat 6 data yang salah diklasifikasikan sebagai Baik Sekali (FP) dan 2 data yang seharusnya masuk kategori ini tetapi tidak terdeteksi (FN). Jumlah False Positive yang cukup besar menunjukkan adanya kecenderungan model untuk memilih kelas ini saat prediksi tidak meyakinkan. Hal ini mengindikasikan munculnya middle bias, yakni bias model untuk memilih kelas tengah dalam kondisi ambiguitas, mengingat posisi Baik Sekali yang berada di antara Baik dan Unggul.

Pada kategori Unggul, performa model tampak lebih baik. Terdapat 13 data yang berhasil diklasifikasikan secara benar (TP) dan 20 data lainnya yang diidentifikasi dengan tepat sebagai bukan Unggul (TN). Namun demikian, masih terdapat 1 data yang salah diklasifikasikan sebagai Unggul padahal bukan (FP), serta 2 data yang seharusnya masuk ke kategori ini namun tidak dikenali (FN). Meskipun kesalahan masih ada, performa model pada kategori ini relatif stabil, menunjukkan bahwa *Logistic Regression* cukup mampu mengenali pola karakteristik mutu Unggul.



Gambar 5.17 Skor AUC model LR-C

Selanjutnya pada model terbaik *Logistic Regression* juga dilakukan analisis lanjutan terbaik menggunakan analisis Kurva ROC (*Receiver Operating Characteristic*). Visualisasi Kurva ROC dari model **LR-C** dalam penelitian ini yang ditampilkan pada Gambar 5.17 memperoleh skor AUC (*Area Under Curve*) untuk kelas ‘Unggul’ sebesar 0.98, kelas ‘Baik Sekali’ sebesar 0.91, dan untuk

kelas ‘Baik’ sebesar 0.93. Skor AUC yang mendekati 1.0 ini mengindikasikan bahwa model memiliki performa yang tinggi untuk dapat membedakan antara perguruan tinggi di Indonesia yang memiliki akreditasi ‘Unggul’, ‘Baik Sekali’, dan ‘Baik’.

5.4. Kesimpulan

Berdasarkan hasil evaluasi terhadap lima skenario model *Logistic Regression* (LR-A hingga LR-E), dapat disimpulkan bahwa model ini menunjukkan performa klasifikasi yang stabil, efisien, dan cukup kompetitif dalam mengukur kualitas perguruan tinggi di Indonesia. Di antara kelima model, LR-C menunjukkan performa terbaik dengan nilai akurasi sebesar 77,78%, serta didukung oleh precision, recall, dan f1-score yang lebih merata dibandingkan model lainnya. Hal ini menunjukkan bahwa LR-C mampu menangkap pola data secara lebih menyeluruh dan seimbang antar kategori klasifikasi mutu perguruan tinggi, yakni Baik, Baik Sekali, dan Unggul.

Model LR-A, LR-B, dan LR-D mencatatkan akurasi serupa sebesar 75%, namun dengan ketidakseimbangan metrik di beberapa kategori, terutama pada kategori “Baik” yang cenderung memiliki nilai recall yang lebih rendah. Sementara itu, LR-E menunjukkan performa paling rendah dengan akurasi 70%, disertai nilai precision dan recall yang lebih fluktuatif, menunjukkan adanya kelemahan dalam mengklasifikasikan data secara konsisten, terutama ketika proporsi data latih dan uji dibagi secara merata.

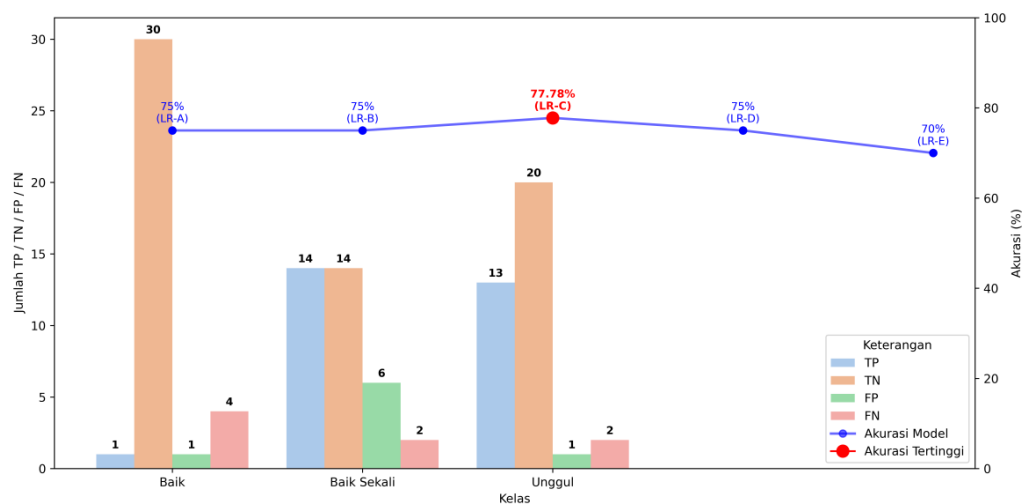
Berdasarkan analisis nilai TP, TN, FP, dan FN dari model terbaik (LR-C),

ditemukan bahwa meskipun model menunjukkan akurasi tinggi secara umum, terdapat ketidakseimbangan dalam klasifikasi antar kelas. Misalnya, model cenderung lebih akurat dalam mengidentifikasi kategori Unggul (dengan TP dan TN yang tinggi serta FP rendah), namun kurang sensitif terhadap kategori “Baik”, yang terlihat dari jumlah FN yang lebih tinggi. Kondisi ini mengindikasikan adanya middle bias, yakni kecenderungan model untuk memilih kategori menengah (Baik Sekali) saat menghadapi ambiguitas, yang umum terjadi pada klasifikasi ordinal.

Analisis lanjutan melalui kurva ROC pada model LR-C semakin menguatkan kesimpulan ini. Nilai AUC untuk masing-masing kelas menunjukkan performa pemisahan yang sangat baik yaitu 0.98 untuk kelas “Unggul”, 0.91 untuk kelas “Baik Sekali”, dan 0.93 untuk kelas Baik. Nilai AUC yang mendekati 1.0 mengindikasikan bahwa model LR-C memiliki kemampuan tinggi dalam membedakan antar kelas, meskipun tetap diperlukan penyesuaian threshold atau teknik seperti cost-sensitive learning untuk mengatasi ketidakseimbangan sensitivitas di masing-masing kelas.

Temuan dalam penelitian ini, yaitu di tunjukkan pada Gambar 5.18 khususnya pada model *Logistic Regression*. Pada model **LR-C** yang menunjukkan akurasi tinggi (77,78%) dan skor AUC mendekati 1.0, sejalan dengan penelitian terdahulu yang menyoroti keandalan *Logistic Regression* dalam tugas klasifikasi multi-kelas yang bersifat ordinal. Pada penelitian yang dilakukan oleh oleh Choudhury & Nur, (2019) menunjukkan bahwa *Logistic Regression* dapat memberikan performa yang stabil dan interpretatif pada dataset klasifikasi

pendidikan, terutama ketika distribusi antar kelas tidak terlalu timpang. Penelitian ini juga mencatat bahwa metode ini unggul dalam efisiensi dan kesederhanaan interpretasi dibandingkan model non-linier seperti *Random Forest* atau Neural Network, terutama pada dataset berskala kecil hingga menengah.



Gambar 5.18 Perbandingan performa model *Logistic Regression*

Selain itu, studi oleh Mitra, (2023) mengenai prediksi akreditasi institusi pendidikan tinggi menggunakan algoritma klasifikasi juga menemukan bahwa *Logistic Regression* memiliki kemampuan prediktif yang baik, namun performanya sangat dipengaruhi oleh distribusi kelas dan kualitas fitur. Mereka menekankan pentingnya teknik penyeimbangan data dan eksplorasi fitur tambahan untuk memaksimalkan performa model, yang juga tercermin dalam temuan penelitian ini, terutama pada model LR-E yang mengalami penurunan performa saat distribusi data tidak optimal.

Di sisi lain, temuan mengenai adanya middle bias dalam model LR-C, di mana model lebih sering memilih kategori menengah ("Baik Sekali") saat menghadapi ambiguitas klasifikasi, merupakan fenomena yang juga ditemukan

dalam berbagai penelitian klasifikasi ordinal. Middle bias biasanya muncul ketika perbedaan antara kelas berurutan tidak cukup tajam atau fitur prediktif tidak cukup kuat membedakan antar kelas, seperti diungkapkan oleh Hernández-Lobato *et al.*, (2014) dalam studi mereka tentang ordinal regression.

Penelitian oleh N. Li & Jimenez, (2018) juga menunjukkan bahwa model klasifikasi berbasis probabilistik, seperti *Logistic Regression*, cenderung menempatkan prediksi di kelas tengah ketika distribusi kelas saling tumpang tindih dan batas pengambilan keputusan tidak cukup tegas. Hal ini semakin diperkuat oleh studi Leevy *et al.*, (2018) yang mengamati bahwa middle bias bisa muncul karena kurangnya data pembeda di kelas ekstrem, seperti kelas “Baik” dalam penelitian ini yang memiliki lebih banyak false negative, sehingga model gagal mengenalinya secara akurat.

BAB VI

PEMBAHASAN

Pada satu dekade terakhir, sistem penilaian mutu perguruan tinggi di Indonesia masih sangat mengandalkan pendekatan administratif yang bersifat formal, seperti proses akreditasi dan penyusunan dokumen pendukung. Meskipun akreditasi berperan penting dalam menjamin terpenuhinya standar minimal sebuah institusi, proses ini cenderung berlangsung secara berkala dan kurang mencerminkan perubahan kualitas yang terjadi secara cepat di lapangan. Sering kali, perguruan tinggi yang sudah melakukan banyak pembaruan dan peningkatan layanan belum memperoleh pengakuan langsung melalui sistem penilaian yang ada. Alhasil, ada kesenjangan antara kualitas riil institusi dengan hasil evaluasi administratif yang diterima.

Ketika penilaian hanya bertumpu pada dokumen, maka realitas di lapangan bisa jadi luput dari perhatian. Evaluasi berbasis laporan tertulis hanya menangkap sisi yang terlihat dan dilaporkan, bukan keseluruhan proses yang terjadi dalam kehidupan kampus secara nyata. Akibatnya, capaian penting seperti peningkatan produktivitas riset, keterlibatan mahasiswa dalam kegiatan internasional, hingga kemitraan strategis dengan dunia industri, sering kali tidak tercermin secara utuh dalam penilaian kualitas. Dengan pendekatan seperti ini, institusi yang aktif dan dinamis justru bisa terabaikan karena kurang terwakili oleh indikator penilaian yang terlalu administratif.

Tantangan semakin terasa ketika kita menyadari bahwa dunia pendidikan tinggi bergerak begitu cepat, terlebih di era digital seperti saat ini. Perguruan

tinggi dituntut untuk terus berinovasi baik dalam hal kurikulum, teknologi pembelajaran, hingga pengembangan kompetensi dosen dan mahasiswa. Sayangnya, sistem evaluasi yang ada belum mampu mengikuti irama perubahan tersebut. Karena sifatnya yang lambat dan konvensional, pendekatan administratif tidak cukup fleksibel untuk menangkap dinamika mutu yang terus berubah. Akibatnya, ada kesenjangan antara kondisi nyata institusi dan cara institusi itu dinilai secara resmi oleh publik maupun lembaga pengawas.

Di tengah keterbatasan metode evaluasi tradisional, muncul kebutuhan akan pendekatan yang lebih adaptif dan berbasis data untuk mengukur kualitas perguruan tinggi secara lebih objektif dan real-time. Salah satu solusi yang kini mulai mendapat perhatian adalah penerapan *Machine Learning*. Teknologi ini menawarkan cara baru dalam memproses informasi pendidikan yang jumlahnya besar dan kompleks, dengan kemampuan mengenali pola-pola tersembunyi yang sering kali tidak terlihat melalui analisis konvensional. Dalam konteks inilah, *Machine Learning* menjadi alat bantu yang menjanjikan untuk memperkuat sistem evaluasi pendidikan tinggi di Indonesia.

Dua model yang populer dan sering digunakan dalam klasifikasi berbasis data adalah *Random Forest* dan *Logistic Regression*. Kedua metode ini tidak hanya mampu bekerja dengan cepat dan efisien, tetapi juga handal dalam menangani variabel yang banyak dan bersifat heterogen. *Random Forest*, misalnya, dikenal mampu membangun sejumlah besar pohon keputusan yang kemudian digabungkan untuk menghasilkan prediksi yang kuat dan stabil. Di sisi lain, *Logistic Regression* unggul dalam menyederhanakan hubungan antara

variabel input dan output, sehingga cocok untuk interpretasi yang jelas dan akurat. Kombinasi keduanya memungkinkan proses penilaian mutu menjadi lebih komprehensif.

Penerapan model *Machine Learning* seperti ini sangat relevan dalam ekosistem pendidikan tinggi yang dinamis. Dengan bantuan algoritma cerdas, evaluasi tidak lagi hanya bergantung pada laporan tertulis, tetapi dapat mencerminkan performa institusi secara lebih menyeluruh mulai dari capaian akademik, kontribusi riset, hingga pengaruh sosial dari sebuah perguruan tinggi. Selain itu, *Machine Learning* juga mampu terus belajar dan menyesuaikan diri dengan data baru, sehingga sistem evaluasi bisa menjadi lebih responsif terhadap perubahan kualitas secara berkelanjutan. Hal ini membuka peluang bagi terciptanya sistem penilaian mutu pendidikan yang lebih adil, transparan, dan berbasis kinerja aktual.

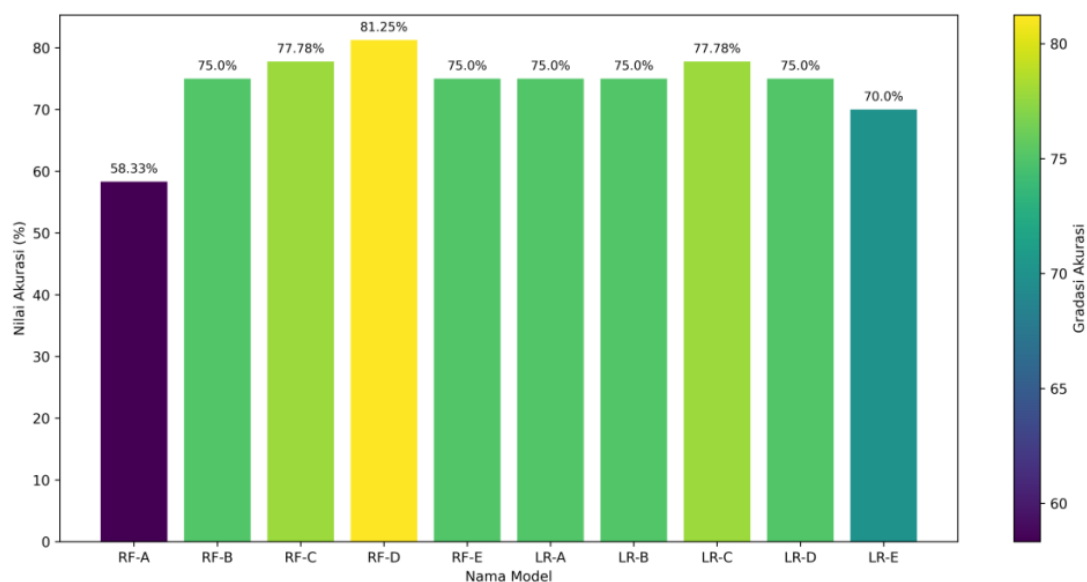
Rumusan masalah dalam penelitian ini disusun dengan mengacu pada dua pertanyaan kunci yang mendasar. Pertama, bagaimana strategi paling tepat dalam mengukur kualitas perguruan tinggi dengan pendekatan *Machine Learning*, dan kedua, seberapa tinggi tingkat akurasi dari masing-masing model yang digunakan, yaitu *Random Forest* dan *Logistic Regression*. Dengan merangkum aspek strategi dan performa teknis, penelitian ini tidak hanya menitikberatkan pada efektivitas algoritma, tetapi juga mempertimbangkan relevansi dan kelayakan implementasinya dalam sistem penilaian mutu pendidikan tinggi. Pendekatan ini memberikan ruang analisis yang lebih luas dan mendalam.

Melalui rumusan tersebut, tujuan utama dari penelitian ini menjadi jelas, yaitu mengevaluasi serta membandingkan performa dua model klasifikasi populer *Random Forest* dan *Logistic Regression* dalam konteks pengukuran kualitas institusi pendidikan tinggi di Indonesia. Evaluasi ini dilakukan untuk memahami kelebihan dan keterbatasan masing-masing model dalam menangkap pola mutu dari data yang tersedia. Harapannya, hasil penelitian ini dapat memberikan masukan yang konkret dan berbasis data kepada pihak-pihak yang terlibat dalam pengambilan kebijakan, khususnya dalam merancang sistem penilaian mutu yang lebih adil, responsif terhadap perkembangan zaman, dan mampu mencerminkan kualitas institusi secara lebih akurat.

Pada tahap awal penelitian, dilakukan serangkaian pengujian terhadap lima skenario rasio data untuk masing-masing model klasifikasi, yaitu *Logistic Regression* (LR-A hingga LR-E) dan *Random Forest* (RF-A hingga RF-E). Setiap skenario menggunakan proporsi data latih dan data uji yang berbeda, seperti 90:10, 80:20, hingga 60:40. Langkah ini dilakukan untuk mengeksplorasi sejauh mana komposisi data latih dan uji dapat mempengaruhi performa model dalam melakukan klasifikasi mutu perguruan tinggi. Dengan merancang pengujian berdasarkan variasi data, penelitian ini mencoba menggambarkan realitas penerapan model dalam berbagai skenario implementasi di lapangan, di mana distribusi data tidak selalu seragam.

Pada Gambar 6.1 menunjukkan hasil pengujian menunjukkan bahwa performa model *Random Forest* cenderung dipengaruhi oleh rasio data yang digunakan. Model **RF-D**, yang menggunakan rasio 60% data latih dan 40% data

uji, berhasil mencatatkan akurasi tertinggi yaitu 81,25%. Temuan ini menunjukkan bahwa keseimbangan antara data pelatihan dan pengujian dapat membantu model dalam menangkap pola dengan lebih baik dan mencegah overfitting. Dalam hal ini, proporsi yang tidak terlalu dominan pada data latih memberikan ruang bagi model untuk diuji secara lebih ketat, sehingga kemampuannya dalam melakukan prediksi terhadap data baru menjadi lebih teruji. Hal ini memberikan wawasan penting bagi praktik pengembangan model klasifikasi yang lebih robust dan dapat diandalkan.



Gambar 6. 1 Perbandingan akurasi model *Random Forest* dan *Logistic Regression*

Sementara itu, model *Logistic Regression* menunjukkan kinerja yang relatif lebih konsisten di berbagai skenario. Di antara kelima varian model yang diuji, **LR-C** dengan rasio data latih dan uji sebesar 70:30 muncul sebagai model dengan performa terbaik. Dengan nilai akurasi mencapai 77,78%, model ini menunjukkan kemampuan klasifikasi yang solid dan cukup dapat diandalkan, terutama dalam konteks distribusi data yang tidak terlalu ekstrem. Stabilitas performa ini

menunjukkan bahwa *Logistic Regression* memiliki daya tahan yang baik terhadap variasi rasio data, menjadikannya sebagai pilihan yang relevan untuk diterapkan pada sistem penilaian mutu institusi pendidikan tinggi, terutama ketika efisiensi dan kejelasan interpretasi hasil menjadi pertimbangan utama.

Tabel 6. 1 Performa kinerja model terbaik (RF-D & LR-C)

Model	Kelas	Precision (%)	Recall (%)	F1-Score (%)
<i>Random Forest</i>	Baik	83	71	77
	Baik Sekali	75	86	80
	Unggul	89	80	84
<i>Logistic Regresion</i>	Baik	50	20	29
	Baik Sekali	70	88	78
	Unggul	93	87	90

Lebih lanjut, pada Tabel 6.1 menunjukkan hasil analisis terhadap metrik evaluasi seperti precision, recall, dan f1-score turut memperkuat pemahaman terhadap performa masing-masing model. *Random Forest* khususnya model RF-D menunjukkan dominasi pada nilai recall di hampir semua kategori mutu perguruan tinggi. Hal ini berarti bahwa model ini memiliki kemampuan yang lebih baik dalam mengenali seluruh data yang termasuk dalam suatu kategori, atau dengan kata lain, sensitif terhadap keberadaan kelas target. Kemampuan ini sangat krusial dalam konteks pengukuran kualitas, karena institusi yang sebenarnya memenuhi kriteria mutu tertentu tidak akan terlewat dalam proses klasifikasi. Dengan demikian, RF-D unggul dalam cakupan prediksi, sedangkan LR-C menonjol dalam kestabilan dan efisiensi klasifikasi.

Di sisi lain, model LR-C menunjukkan keseimbangan kinerja yang mengesankan antara precision dan recall, terutama dalam mengklasifikasikan kategori "Unggul." Kemampuan ini mengindikasikan bahwa model tidak hanya akurat dalam memberikan prediksi yang tepat, tetapi juga mampu menangkap

sebagian besar institusi yang benar-benar termasuk dalam kategori tersebut. Meskipun *Logistic Regression* memiliki struktur yang lebih sederhana dibandingkan *Random Forest*, hasil ini menunjukkan bahwa kesederhanaan tidak selalu berarti kelemahan. Justru, pada konteks tertentu, seperti saat interpretabilitas dan efisiensi komputasi dibutuhkan, model ini dapat menjadi alternatif yang kompetitif. Keseimbangan antara efektivitas dan efisiensi yang dimiliki LR-C membuatnya layak dipertimbangkan dalam sistem evaluasi mutu pendidikan tinggi yang adil dan berkelanjutan.

Tabel 6. 2 Nilai TP, TN, FP, FN model terbaik (RF-D & LR-C)

Model	Kelas	TP	TN	FP	FN
<i>Random Forest</i>	Baik	5	40	1	2
	Baik Sekali	18	21	6	3
	Unggul	16	26	2	4
<i>Logistic Regresion</i>	Baik	1	30	1	4
	Baik Sekali	14	14	6	2
	Unggul	13	20	1	2

Pada Tabel 6.2 menunjukkan terkait dengan hasil analisis lebih mendalam terhadap nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) pada model RF-D menunjukkan kecenderungan model ini untuk bersikap konservatif dalam proses klasifikasi. Model ini menghasilkan jumlah FP yang rendah, yang berarti sangat jarang salah menempatkan institusi dalam kategori mutu yang lebih tinggi dari seharusnya. Di sisi lain, RF-D tetap mampu menjaga tingkat TP yang tinggi, yang mencerminkan ketepatan dalam mengenali institusi yang memang layak mendapatkan kategori tertentu. Karakteristik ini sangat sesuai untuk digunakan dalam sistem evaluasi mutu yang membutuhkan ketelitian dan akurasi tinggi, seperti proses akreditasi atau pemberian penghargaan kualitas.

Pendekatan konservatif RF-D juga menjadi nilai tambah dalam konteks pengambilan keputusan yang berdampak besar, seperti alokasi anggaran pendidikan atau perumusan kebijakan peningkatan mutu. Kesalahan dalam mengklasifikasikan perguruan tinggi ke dalam kategori yang lebih tinggi bisa berakibat pada ketimpangan sumber daya dan persepsi publik. Dengan menghasilkan sedikit FP, RF-D membantu menjaga keadilan dalam sistem penilaian, memastikan bahwa hanya institusi yang benar-benar memenuhi kriteria yang diakui secara formal. Ini penting dalam menjaga kepercayaan terhadap sistem evaluasi mutu nasional.

Nilai Akurasi yang cukup tinggi secara keseluruhan, terdapat fenomena yang disebut dengan middle bias. Middle bias ini merujuk pada kecenderungan model untuk mengklasifikasikan data ke kategori menengah, yakni “Baik Sekali”, terutama ketika model mengalami ketidakpastian dalam memutuskan apakah data tersebut seharusnya masuk kategori "Baik" atau "Unggul". Akibatnya, model bisa kehilangan sensitivitas terhadap data yang berada di dua ujung spektrum mutu, yang tentu dapat memengaruhi ketepatan evaluasi.

Fenomena middle bias pada LR-C patut diperhatikan secara serius, khususnya dalam konteks klasifikasi ordinal seperti penilaian kualitas perguruan tinggi. Jika tidak ditangani, bias ini dapat menyebabkan institusi yang seharusnya mendapatkan pengakuan sebagai "Unggul" justru disamakan dengan yang "Baik Sekali", atau sebaliknya. Hal ini tentu mengaburkan gambaran mutu yang sebenarnya dan bisa berdampak pada pengambilan keputusan yang kurang tepat. Untuk mengatasi masalah ini, strategi seperti penyesuaian threshold prediksi atau

penerapan pembobotan kelas bisa menjadi solusi yang efektif agar sensitivitas model terhadap masing-masing kategori dapat ditingkatkan.

Tabel 6. 3 Skor AUC pada performa model terbaik (RF-D & LR-C)

Kelas	Model	Skor AUC
Baik	<i>Random Forest</i>	0.90
	<i>Logistic Regresion</i>	0.93
Baik Sekali	<i>Random Forest</i>	0.81
	<i>Logistic Regresion</i>	0.91
Unggul	<i>Random Forest</i>	0.92
	<i>Logistic Regresion</i>	0.98

Pada Tabel 6.3 menunjukkan bahwa middle bias yang ditemukan pada model LR-C menjadi perhatian penting, mengingat klasifikasi mutu perguruan tinggi bersifat ordinal. Kesalahan prediksi antar kategori, seperti mengklasifikasikan institusi "Unggul" menjadi "Baik Sekali", dapat menyebabkan interpretasi yang keliru dan berisiko terhadap reputasi serta kebijakan yang diambil. Meskipun akurasi tinggi, model ini cenderung menetralkan data ke kelas menengah. Oleh karena itu, diperlukan penyesuaian threshold atau strategi pembobotan agar prediksi lebih sensitif terhadap perbedaan antar tingkat mutu. Namun demikian, performa LR-C tetap menunjukkan hasil yang sangat baik, sebagaimana diperkuat oleh analisis ROC Curve. Skor AUC yang tinggi untuk semua kategori 0,98 untuk "Unggul", 0,91 untuk "Baik Sekali", dan 0,93 untuk "Baik" menandakan kemampuan model dalam membedakan kategori secara akurat. Hal ini menunjukkan bahwa *Logistic Regression* tetap kompetitif sebagai metode klasifikasi mutu perguruan tinggi, terlebih karena model ini juga lebih sederhana dan mudah diinterpretasikan dibanding algoritma yang lebih kompleks seperti *Random Forest*.

Model *Random Forest* juga menunjukkan performa yang cukup kompetitif

dalam analisis AUC (Area Under Curve), meskipun nilai AUC-nya sedikit berada di bawah model LR-C. Temuan ini memperkuat gagasan bahwa kedua algoritma baik *Random Forest* maupun *Logistic Regression* memiliki keunggulan masing-masing yang bisa dioptimalkan sesuai dengan tujuan evaluasi. *Random Forest*, misalnya, unggul dalam menangani data dengan pola kompleks dan fitur yang banyak, sementara *Logistic Regression* unggul dari sisi interpretabilitas dan stabilitas hasil. Performa AUC yang layak pada *Random Forest* menunjukkan bahwa model ini masih mampu melakukan klasifikasi dengan baik dan membedakan antar kategori mutu secara akurat.

Dalam konteks penilaian kualitas perguruan tinggi yang semakin kompleks dan dinamis, tidak cukup hanya bergantung pada metrik akurasi secara keseluruhan. Model klasifikasi yang ideal adalah model yang mampu menjaga keseimbangan antara sensitivitas (kemampuan mengenali kategori dengan benar) dan spesifisitas (kemampuan menghindari kesalahan klasifikasi). Hal ini penting agar hasil klasifikasi tidak condong ke satu kategori saja, melainkan mampu merepresentasikan distribusi mutu secara adil dan proporsional. Oleh karena itu, pemilihan model tidak bisa semata-mata didasarkan pada angka akurasi tertinggi, tetapi juga mempertimbangkan distribusi prediksi yang seimbang, relevansi konteks, serta dampak dari keputusan klasifikatif tersebut terhadap institusi dan pemangku kebijakan.

Model RF-D menunjukkan keunggulan dari sisi akurasi total, yang berarti model ini mampu memprediksi kualitas perguruan tinggi dengan ketepatan yang tinggi dibanding model lainnya. Namun, keunggulan ini datang dengan

konsekuensi penggunaan sumber daya komputasi yang lebih besar, baik dari segi waktu pelatihan maupun kebutuhan memori. Di sisi lain, *Logistic Regression*, khususnya model LR-C, menawarkan kelebihan dalam hal efisiensi dan kestabilan. Struktur modelnya yang lebih sederhana membuat proses pelatihan lebih cepat dan hasilnya lebih mudah diinterpretasikan. *Logistic Regression* juga memiliki risiko overfitting yang lebih rendah, menjadikannya pilihan yang ideal dalam skenario evaluasi yang menuntut transparansi dan efisiensi tinggi.

Temuan ini konsisten dengan hasil dari beberapa penelitian sebelumnya. (Latifah & Anggitha, n.d.) mencatat bahwa *Random Forest* memiliki performa unggul dalam menangani data dengan struktur kompleks dan hubungan non-linear antar variabel. Sebaliknya, *Logistic Regression* lebih optimal digunakan pada data dengan distribusi kelas yang seimbang dan pola hubungan antar variabel yang linier. Penelitian lain oleh (Rahmawati, n.d.) juga mendukung keunggulan *Logistic Regression* di sektor pendidikan, khususnya karena kemampuan model ini dalam menyajikan hasil yang mudah dipahami oleh pengambil kebijakan. Hal ini menjadi penting dalam proses pengambilan keputusan strategis, terutama dalam penilaian mutu institusi pendidikan yang memiliki implikasi besar terhadap perencanaan dan pendanaan.

Berdasarkan data yang dianalisis dalam penelitian ini, dapat disimpulkan bahwa variabel-variabel administratif seperti rasio dosen-mahasiswa, status akreditasi program studi, jumlah publikasi jurnal terakreditasi, serta intensitas kegiatan mahasiswa memiliki peran penting dalam membedakan mutu suatu perguruan tinggi. Meskipun sering kali dianggap sebagai indikator formal, fitur-

fitur ini ternyata mampu merepresentasikan berbagai aspek kualitas institusi secara cukup efektif ketika dikombinasikan dalam model prediktif. Keberadaan variabel-variabel tersebut membantu algoritma dalam mengenali pola tertentu yang terkait erat dengan klasifikasi mutu institusi, baik dari sisi input sumber daya maupun output akademik dan non-akademik.

Lebih lanjut, fitur-fitur tersebut menunjukkan kontribusi yang signifikan terhadap performa model, khususnya bila dioptimalkan melalui pendekatan seleksi fitur yang tepat. Teknik seperti recursive feature elimination (RFE) atau analisis feature importance yang tersedia dalam algoritma *Random Forest*, dapat dimanfaatkan untuk menyaring variabel yang paling relevan dan menghilangkan fitur yang redundan atau tidak informatif. Dengan demikian, model yang dibangun tidak hanya menjadi lebih efisien secara komputasional, tetapi juga lebih akurat dan interpretatif. Pendekatan ini penting dalam konteks evaluasi mutu pendidikan tinggi, karena memungkinkan fokus terhadap faktor-faktor yang benar-benar berdampak terhadap peningkatan kualitas institusi.

Dalam upaya mengembangkan sistem penilaian mutu pendidikan tinggi yang lebih adaptif dan responsif terhadap dinamika aktual, pemanfaatan model *Machine Learning* seperti *Random Forest* dan *Logistic Regression* dapat menjadi fondasi penting dalam membangun dashboard evaluasi berbasis data secara real time. Model-model ini mampu memproses data dalam jumlah besar dengan kecepatan tinggi, sekaligus mengidentifikasi pola tersembunyi yang sulit ditangkap oleh metode evaluasi konvensional. Dengan pendekatan ini, pemantauan kualitas institusi tidak lagi bersifat statis atau periodik semata,

melainkan dapat dilakukan secara terus-menerus dan otomatis, mengikuti perubahan yang terjadi di lapangan.

Di sisi lain, salah satu keunggulan utama dari model *Logistic Regression* yang tidak boleh diabaikan adalah tingkat interpretabilitasnya yang tinggi. Hal ini menjadikannya sangat relevan dalam konteks penilaian oleh lembaga akreditasi atau regulator pendidikan yang menuntut transparansi, akuntabilitas, dan kejelasan dalam proses pengambilan keputusan. Karena model ini menghasilkan parameter yang dapat ditafsirkan secara langsung, pemangku kebijakan dapat memahami alasan di balik klasifikasi mutu suatu perguruan tinggi sehingga hasil evaluasi tidak hanya akurat, tetapi juga dapat dipertanggungjawabkan secara ilmiah maupun administratif.

Penyesuaian threshold prediksi merupakan salah satu strategi teknis yang cukup efektif dalam mengatasi permasalahan middle bias, khususnya pada model *Logistic Regression*. Dalam konteks klasifikasi ordinal seperti mutu perguruan tinggi, di mana ketepatan prediksi terhadap kelas ekstrem (misalnya “Unggul” atau “Baik”) sangat penting, threshold yang terlalu netral atau default dapat menyebabkan prediksi cenderung menumpuk di kelas tengah. Dengan mengkalibrasi ambang batas prediksi berdasarkan distribusi data atau melalui teknik seperti ROC analysis, model dapat diarahkan untuk membuat keputusan klasifikasi yang lebih presisi dan mencerminkan realitas yang lebih akurat dari kualitas institusi yang dievaluasi.

Lebih jauh lagi, penelitian ini memberikan gambaran bahwa pendekatan hybrid yakni dengan menggabungkan kekuatan *Random Forest* dalam mendeteksi

pola non-linear dan kompleksitas interaksi fitur dengan keunggulan interpretabilitas *Logistic Regression* dapat menjadi solusi ideal dalam sistem penilaian pendidikan tinggi. Sistem semacam ini mampu memadukan akurasi tinggi dan transparansi hasil, sehingga tidak hanya mendukung pengambilan keputusan berbasis data, tetapi juga dapat meningkatkan kepercayaan publik dan akuntabilitas dalam proses evaluasi institusi pendidikan. Pendekatan hybrid ini dinilai sangat relevan dalam era big data dan pendidikan berbasis digital, di mana kebutuhan akan sistem penilaian yang cerdas, adaptif, dan efisien menjadi semakin mendesak.

Selain mempertimbangkan performa teknis seperti akurasi, presisi, dan recall, penerapan model prediksi mutu pendidikan berbasis *Machine Learning* juga harus memperhatikan aspek etika serta dampak kebijakan yang mungkin timbul. Keputusan yang dihasilkan oleh model memiliki potensi untuk memengaruhi reputasi institusi, distribusi sumber daya, bahkan kepercayaan publik terhadap sistem pendidikan. Oleh karena itu, penting untuk memastikan bahwa model tidak memperkuat bias historis atau menciptakan ketimpangan baru. Misalnya, institusi yang secara historis memiliki keterbatasan sumber daya bisa saja diprediksi memiliki mutu rendah, padahal sebenarnya mereka sedang dalam fase peningkatan kualitas yang belum tercermin penuh dalam data.

Dari sudut pandang implementasi praktis, *Logistic Regression* menjadi pilihan yang lebih bijak untuk tahap awal pengembangan sistem karena kemudahan dalam interpretasi dan integrasinya dengan sistem evaluasi yang sudah ada. Model ini memberikan transparansi dalam logika pengambilan

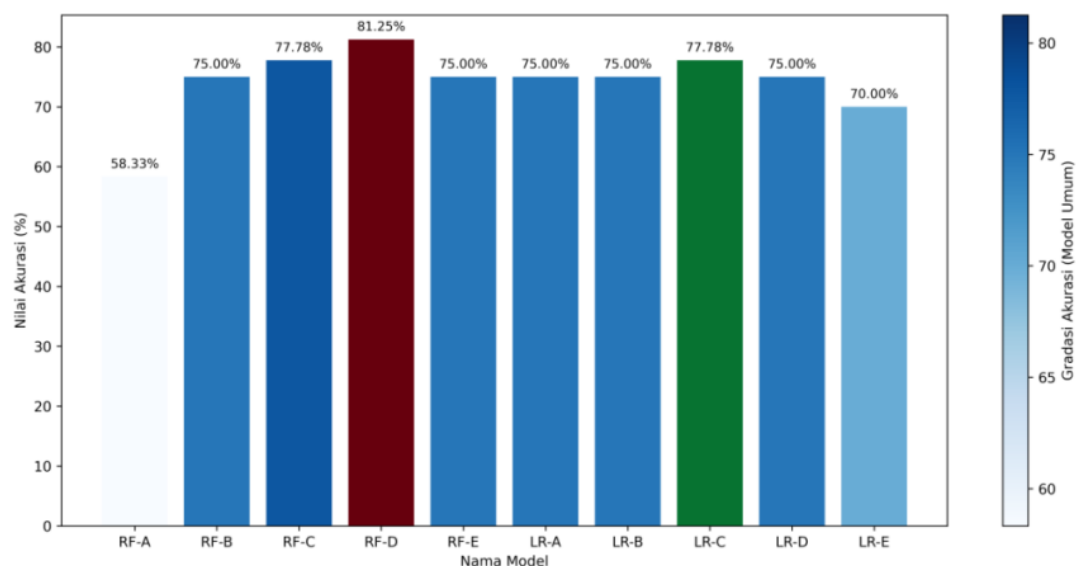
keputusan, sehingga lebih mudah dipahami oleh pemangku kebijakan, pengelola institusi, dan publik. Sementara itu, *Random Forest* dapat berperan sebagai alat validasi silang (cross-validation) karena kemampuannya menangkap hubungan kompleks antar variabel yang mungkin terlewatkan oleh model linier. Kombinasi keduanya menciptakan kerangka evaluasi yang kuat, fleksibel, dan minim risiko dari sisi etika maupun teknis.

Kombinasi antara *Random Forest* dan *Logistic Regression* memberikan landasan yang lebih kokoh dalam proses evaluasi mutu perguruan tinggi. Pendekatan ini memungkinkan model yang tidak hanya andal secara statistik, tetapi juga dapat disesuaikan dengan kebutuhan pengambilan keputusan di dunia nyata. *Random Forest* memberikan kekuatan dalam mendeteksi pola-pola kompleks dari data yang besar dan beragam, sementara *Logistic Regression* menawarkan interpretabilitas tinggi yang sangat dibutuhkan dalam konteks kebijakan publik. Dengan demikian, evaluasi mutu pendidikan tinggi tidak lagi sekadar berbasis angka, tetapi juga mampu menjelaskan alasan di balik klasifikasi yang dihasilkan, yang penting untuk perbaikan berkelanjutan.

Oleh sebab itu, penentuan model terbaik tidak bisa hanya didasarkan pada capaian akurasi tertinggi semata. Efisiensi dalam penggunaan sumber daya, kemampuan menjelaskan hasil (explainability), serta keseimbangan performa antar kategori menjadi faktor penentu yang tidak kalah penting. Model yang akurat namun sulit diinterpretasikan dapat menyulitkan lembaga pendidikan dalam merespons hasil evaluasi. Sebaliknya, model yang transparan namun kurang sensitif terhadap dinamika data juga memiliki keterbatasan. Maka,

penggabungan kekuatan dua model ini menciptakan pendekatan yang tidak hanya teknis unggul, tetapi juga visioner dalam mendorong peningkatan mutu pendidikan tinggi secara berkelanjutan dan inklusif.

Berdasarkan hasil evaluasi menyeluruh terhadap lima skenario dari masing-masing model, dapat disimpulkan bahwa model **LR-C** merupakan pilihan utama yang layak direkomendasikan dalam konteks pengukuran mutu perguruan tinggi di Indonesia. Model ini menampilkan performa klasifikasi yang konsisten, efisien, serta memiliki interpretabilitas tinggi sebuah aspek yang krusial dalam dunia pendidikan tinggi, di mana hasil klasifikasi harus bisa dipahami dan ditindaklanjuti oleh pemangku kebijakan. Keunggulan LR-C dalam menyeimbangkan antara precision dan recall, serta skor AUC yang mendekati sempurna di setiap kelas, menjadikannya solusi yang praktis dan handal.



Gambar 6. 2 Performa kinerja model terbaik

Sementara itu, model **RF-D** yang mencatatkan akurasi tertinggi menunjukkan potensi besar dalam konteks validasi silang atau pengujian

mendalam, terutama ketika digunakan untuk mendeteksi pola-pola tersembunyi dalam data kompleks. Kombinasi keduanya, LR-C untuk sistem utama dan RF-D sebagai alat validasi, dapat menghadirkan pendekatan evaluasi mutu yang tidak hanya kuat secara teknis, tetapi juga inklusif dan adaptif. Maka, penelitian ini menyimpulkan bahwa pemanfaatan *Machine Learning* khususnya model *Logistic Regression* dan *Random Forest* telah terbukti mampu mendorong sistem evaluasi kualitas pendidikan tinggi menjadi lebih objektif, responsif terhadap data aktual, dan selaras dengan kebutuhan pengambilan kebijakan berbasis bukti.

Dalam konteks pengukuran kualitas perguruan tinggi yang dilakukan melalui pendekatan *Machine Learning*, nilai-nilai yang mendasari penelitian ini sejatinya juga sejalan dengan prinsip-prinsip keadilan, objektivitas, dan penilaian berbasis data. Islam sangat menekankan pentingnya menilai dengan adil dan berdasarkan bukti nyata, sebagaimana firman Allah berikut:

﴿يَا أَيُّهَا الَّذِينَ آمَنُوا كُونُوا قَوَّامِينَ بِالْقِسْطِ شُهَدَاءَ لِلَّهِ وَلَوْ عَلَىٰ أَنْفُسِكُمْ أَوِ الْوَالِدِينَ وَالْأَقْرَبِينَ إِن يَكُنْ عَدِيًّا أَوْ فَقِيرًا فَاللَّهُ أَوْلَىٰ بِهِمَا فَلَا تَتَّبِعُوا الْهَوَىٰ أَنْ تَعْدِلُوا وَإِنْ تَلَوَّا أَوْ تَعْرِضُوا فَإِنَّ اللَّهَ كَانَ بِمَا تَعْمَلُونَ خَبِيرًا

"Wahai orang-orang yang beriman! Jadilah kamu orang-orang yang benar-benar menegakkan keadilan, menjadi saksi karena Allah, sekalipun terhadap dirimu sendiri atau ibu bapak dan kaum kerabatmu. Jika ia kaya ataupun miskin, maka Allah lebih tahu kemaslahatannya. Maka janganlah kamu mengikuti hawa nafsu karena ingin menyimpang dari kebenaran. Dan jika kamu memutarbalikkan (kata-kata) atau enggan menjadi saksi, maka sungguh, Allah adalah Maha Mengetahui terhadap apa yang kamu kerjakan" (QS. An-Nisa' /4:135, Kemenag 2019).

Menurut penjelasan M. Quraish Shihab (2002), Alqur'an Surat An-Nisa' ayat 135 menegaskan bahwa orang-orang beriman diperintahkan untuk menjadi penegak keadilan (qawwāmīna bil-qisṭi) dalam seluruh aspek kehidupan. Keadilan

harus ditegakkan meskipun hal tersebut berkaitan dengan diri sendiri, kedua orang tua, atau kerabat dekat. Dengan demikian, hubungan emosional, kedekatan keluarga, kepentingan pribadi, ataupun tekanan sosial tidak boleh mempengaruhi suatu keputusan. Ayat ini juga menegaskan bahwa keadaan ekonomi seseorang, baik kaya maupun miskin, tidak boleh dijadikan alasan untuk memihak secara tidak adil. Keduanya berada dalam kekuasaan Allah, sehingga manusia dituntut untuk menempatkan kebenaran dan objektivitas sebagai landasan utama dalam pengambilan keputusan. Quraish Shihab menjelaskan bahwa ayat ini juga memperingatkan dua bentuk penyimpangan: *talwū* (memutarbalikkan fakta) dan *tu'riḍū* (menghindar dari memberikan kesaksian yang benar). Keduanya merupakan sikap tidak adil yang dilarang keras, dan Allah menutup ayat ini dengan penegasan bahwa Dia Maha Teliti atas segala perbuatan manusia.

Berdasarkan penafsiran tersebut, Ayat ini menekankan pentingnya keadilan dan objektivitas dalam setiap bentuk penilaian, termasuk dalam hal mengevaluasi mutu lembaga pendidikan tinggi. Dengan menggunakan model seperti *Logistic Regression* dan *Random Forest*, penilaian terhadap perguruan tinggi dilakukan berdasarkan data nyata, bukan asumsi atau kepentingan subjektif, sehingga mendekati nilai keadilan yang diidealkan Alqur'an.

BAB VII

KESIMPULAN

7.1. Kesimpulan

Berdasarkan hasil analisis data dan evaluasi model, penelitian ini memberikan dua kesimpulan utama yang menjawab fokus pertanyaan penelitian:

1. Model klasifikasi kualitas perguruan tinggi di Indonesia berhasil dikembangkan menggunakan dua pendekatan *Machine Learning*, yaitu *Random Forest* dan *Logistic Regression*. *Logistic Regression* menunjukkan kinerja yang efisien, mudah diinterpretasikan, dan stabil, terutama pada model **LR-C** dengan rasio data latih-uji 70:30. Sementara itu, *Random Forest* khususnya model **RF-D** dengan rasio 60:40 mampu mengidentifikasi pola yang lebih kompleks dan menghasilkan akurasi yang tinggi. Oleh karena itu, penggunaan *Logistic Regression* sangat cocok untuk implementasi awal yang membutuhkan transparansi dan interpretabilitas, sedangkan *Random Forest* dapat dijadikan sebagai model pembanding atau sistem pendukung keputusan untuk memperkuat ketepatan hasil evaluasi mutu perguruan tinggi.
2. Tingkat akurasi dan performa dari kedua metode menunjukkan hasil yang kompetitif, namun dengan karakteristik berbeda. Model *Random Forest* terbaik (RF-D) mencapai akurasi sebesar 81,25% dengan nilai *recall* yang tinggi di seluruh kelas mutu, menunjukkan kekuatan dalam mendeteksi berbagai kategori mutu secara sensitif. Sementara itu, model *Logistic Regression* terbaik (LR-C) mencatat akurasi sebesar 77,78% dengan AUC tinggi di setiap kelas (Unggul: 0,98; Baik Sekali: 0,91; Baik: 0,93), serta

keseimbangan *precision* dan *recall* yang baik. Hal ini membuktikan bahwa *Logistic Regression* tetap mampu menghasilkan performa klasifikasi yang kuat meskipun menggunakan struktur model yang lebih sederhana. Kedua model saling melengkapi, dan pemanfaatannya dapat disesuaikan dengan kebutuhan spesifik dalam proses evaluasi mutu, baik dari segi efisiensi komputasi, interpretabilitas hasil, maupun sensitivitas terhadap distribusi data.

7.2. Keterbatasan

Penelitian ini memberikan kontribusi dalam menilai kualitas perguruan tinggi di Indonesia menggunakan model *Random Forest* dan *Logistic Regression*, namun masih memiliki beberapa keterbatasan. Data yang digunakan bersifat sekunder dan hanya mencakup indikator administratif, belum mencakup aspek penting seperti kepuasan mahasiswa atau kualitas lulusan. Distribusi data yang tidak merata antar kategori mutu juga menimbulkan bias prediksi, terutama pada model *Logistic Regression*. Selain itu, kedua model bersifat supervised dan belum adaptif terhadap dinamika mutu institusi secara waktu nyata. Ruang lingkup penelitian yang terbatas pada konteks Indonesia juga membatasi generalisasi temuan ke sistem pendidikan di negara lain. Oleh karena itu, studi lanjutan dengan cakupan data dan pendekatan yang lebih luas diperlukan untuk membangun sistem evaluasi pendidikan tinggi yang lebih holistik dan berkelanjutan.

DAFTAR PUSTAKA

- Al-Batah, M.S., Alzboon, M.S., Migdadi, H.S., Alkhasawneh, M. dan Alqaraleh, M. 2024. Advanced landslide detection using *Machine Learning* and remote sensing data. *Data Metadata*, 3(1): 1–12.
- Andriani, A. 2012. Penerapan algoritma C4.5 pada program klasifikasi mahasiswa dropout. *Seminar Nasional Matematika*, 139–147. Semarang: Universitas Negeri Semarang.
- Angkotasan, S. dan Watianan, S. 2021. Faktor-faktor yang mempengaruhi peningkatan mutu pendidikan di kampus STIA Alazka Ambon. *Komunitas: Jurnal Ilmu Sosiologi*, 4(2): 42–50.
- Asmawi, M.R. 2005. Strategi meningkatkan lulusan bermutu di perguruan tinggi. *Makara Human Behavior Studies in Asia*, 9(2): 66–71.
- Bhuvaneswari, S. dan Karthikeyan, S. 2023. A comprehensive research of breast cancer detection using *Machine Learning*, clustering and optimization techniques. *Proceedings of the 2023 International Conference on Data Science and Network Security (ICDSNS)*, 1–6.
- Breiman, L. 2001. *Random Forest*. *Machine Learning*, 45: 5–32. Kluwer Academic Publishers. <https://doi.org/10.1023/A:1010933404324>.
- Choudhury, A.M. dan Nur, K. 2019. A *Machine Learning* approach to identify potential customer based on purchase behavior. *Proceedings of the 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 242–247.
- Claassen, C. 2015. Measuring university quality. *Scientometrics*, 104(3): 793–807. <https://doi.org/10.1007/s11192-015-1584-8>.
- Creswell, J. 2014. Penelitian kualitatif dan desain riset: Memilih di antara lima pendekatan. Terj. A. Lazuardi. Yogyakarta: Pustaka Pelajar. (Karya asli diterbitkan 1998).
- Creswell, J.W. dan Creswell, J.D. 2017. *Research design: Qualitative, quantitative, and mixed methods approaches*. California: Sage Publications.
- Dalton, L.A. 2015. Optimal ROC-based classification and performance analysis under Bayesian uncertainty models. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(4): 719–729.
- Djamba, Y.K. 2002. Social research methods: Qualitative and quantitative approaches. *Teaching Sociology*, 30(3): 380.
- Estrada-Real, A.C. dan Cantu-Ortiz, F.J. 2022. A data analytics approach for university competitiveness: the QS world university rankings. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 16(3): 871–891. <https://doi.org/10.1007/s12008-022-00966-2>.

- Fitrah, M. 2018. Urgensi sistem penjaminan mutu internal terhadap peningkatan mutu perguruan tinggi. *Jurnal Penjaminan Mutu*, 4(1): 76–86.
- Haerani, E., Syafria, F., Lestari, F., Novriyanto, N. dan Marzuki, I. 2023. Classification academic data using *Machine Learning* for decision making process. *Journal of Applied Engineering and Technological Science (JAETS)*, 4(2): 955–968. <https://doi.org/10.37385/jaets.v4i2.1983>.
- Han, J., Kamber, M. dan Pei, J. 2011. *Data mining: Concepts and techniques*. Burlington: Morgan Kaufmann.
- Hernández-Lobato, J.M., Houlsby, N. dan Ghahramani, Z. 2014. Probabilistic matrix factorization with non-random missing data. *Proceedings of the International Conference on Machine Learning*, 1512–1520.
- Hosmer, D.W., Lemeshow, S. dan Sturdivant, R.X. 2013. *Applied Logistic Regression*. Wiley Series in Probability and Statistics (1st ed.). Wiley. <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118548387>.
- Ibañez Ramírez, J.S., Echeverri Salazar, T. dan Castrillón Gómez, O.D. 2022. Predicción de la calidad de vida universitaria a través de minería de datos. *Revista Ingenierías Universidad de Medellín*, 21(40): 1–14. <https://doi.org/10.22395/rium.v21n40a1>.
- Indonesia, P.R. 2006. Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional. Jakarta: Sekretariat Negara.
- Indonesia, R. 2012. Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi. Jakarta: Sekretariat Negara.
- Karim, B.A. 2020. Pendidikan perguruan tinggi era 4.0 dalam pandemi Covid-19 (refleksi sosiologis). *Education and Learning Journal*, 1(2): 102–112.
- Latifah, T. dan Anggitha, G.D. n.d. Implementasi metode *Random Forest*, KNN (K-Nearest Neighbour), Decision Tree classification menggunakan *Machine Learning* untuk stroke prediction.
- Leevy, J.L., Khoshgoftaar, T.M., Bauder, R.A. dan Seliya, N. 2018. A survey on addressing high-class imbalance in big data. *Journal of Big Data*, 5(1): 1–30.
- Li, M. 2023. Construction of higher education teaching quality evaluation model based on improved particle swarm optimization algorithm. 120–124. <https://doi.org/10.1145/3645279.3645301>.
- Li, N. dan Jimenez, R. 2018. A *Logistic Regression* classifier for long-term probabilistic prediction of rock burst hazard. *Natural Hazards*, 90(1): 197–215.
- Li, Y. dan Zhang, H. 2024. Big data technology for teaching quality monitoring and improvement in higher education - joint K-means clustering algorithm and Apriori algorithm. *Systems and Soft Computing*, 6: 200125. <https://doi.org/10.1016/j.sasc.2024.200125>.

- Mitra, M. 2023. Urgensi percepatan penjaminan mutu perguruan tinggi melalui akreditasi perguruan tinggi. *Promis*, 4(2): 91–99.
- Mohan, H.G., Kumar, J. dan Nandish, M. 2024. A hybrid feature engineering mechanism based on *Random Forest* for imbalanced botnet dataset. *Proceedings of the 2024 Fourth International Conference on Multimedia Processing, Communication & Information Technology (MPCIT)*, 140–145.
- Musa, A.B. 2013. Comparative study on classification performance between support vector machine and *Logistic Regression*. *International Journal of Machine Learning and Cybernetics*, 4(1): 13–24.
- Muschelli III, J. 2020. ROC and AUC with a binary predictor: a potentially misleading metric. *Journal of Classification*, 37(3): 696–708.
- Mustakim, Z. dan Kamal, R. 2021. K-means clustering for classifying the quality management of secondary education in Indonesia. *Jurnal Cakrawala Pendidikan*, 40(3): 725–737. <https://doi.org/10.21831/cp.v40i3.40150>.
- Nasution, M.K.M. 2016. Karya ilmiah dosen dan mahasiswa. Medan: Harian Waspada.
- Nulhaqim, S.A., Heryadi, D.H., Pancasilawan, R. dan Ferdryansyah, M. 2016. Peranan perguruan tinggi dalam meningkatkan kualitas pendidikan di Indonesia untuk menghadapi ASEAN Community 2015: Studi kasus Universitas Indonesia, Universitas Padjadjaran, dan Institut Teknologi Bandung. *Share: Social Work Journal*, 6(2): 197.
- Prakash, P., Rashmi Siddalingappa, S.K. dan S.S.G. 2024. A knowledge-based grade prediction system using *Machine Learning* for higher education institutions. *Nanotechnology Perceptions*, 20: 1683–1697. <https://doi.org/10.62441/nano-ntp.vi.3001>.
- Rahmawati, D.S. n.d. Analisis sentimen opini publik di media sosial menggunakan algoritma *Machine Learning*.
- Ramdhan, M. 2021. Metode penelitian. Yogyakarta: Cipta Media Nusantara.
- Riset, D.A.N.T. 2022. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. *Universitas*, 1(1): 2.
- Rohmawati, N., Defiyanti, S. dan Jajuli, M. 2015. Implementasi algoritma K-Means dalam pengklasteran mahasiswa pelamar beasiswa. *Jurnal Ilmiah Teknologi Informasi Terapan*, 1(2).
- Romero, C. dan Ventura, S. 2007. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1): 135–146.
- Singgih, M.L. 2008. Faktor-faktor yang mempengaruhi kualitas pendidikan pada perguruan tinggi. *Prosiding Seminar Nasional Teknoin*.
- Syafii, A., Bahar, B., Shobicah, S. dan Muharam, A. 2023. Pengukuran indeks mutu pendidikan berbasis standar nasional. *Jurnal Multidisiplin Indonesia*, 2(7): 1697–1701.

- Syahrir, N., Ansari, M.I. dan Basir, I. 2023. Pendampingan tata kelola keuangan dan pemasaran berbasis digital dalam meningkatkan penjualan. *Jurnal Abdi Insani*, 10(3): 1639–1646.
- Wiraswendro, P.E. dan Soetanto, H. 2022. Application of *Random Forest* Classifier Algorithm in Indonesian Sign Language System (Sibi) Detection System. *BIT: Jurnal Fakultas Teknologi Informasi Universitas Budi Luhur*, 19(2): 75–81.
- Zhang, H., Xiao, B., Li, J. dan Hou, M. 2021. An improved genetic algorithm and neural network-based evaluation model of classroom teaching quality in colleges and universities. *Wireless Communications and Mobile Computing*, 2021(1): 2602385. <https://doi.org/10.1155/2021/2602385>.
- Zhong, Y. 2023. Evaluation and analysis of teaching quality of university teachers using *Machine Learning* algorithms. *Journal of Intelligent Systems*, 32(1): 20220204. <https://doi.org/10.1515/jisys-2022-0204>
- Shihab, M.Q. (2002). *Tafsir Al-Mishbah: Pesan, Kesan, dan Keserasian Al-Qur'an*, Vol. 12. Ciputat: Lentera Hati.