

LAPORAN APPLIED MACHINE LEARNING

**“Analisis Gender dan Prediksi Risiko Putus Sekolah Siswa SMA/MA
Kota Makassar dengan Algoritma Machine Learning
Menggunakan Data 2021–2024 ”**

Dosen Pengampuh : Runal Rezkiawan S.Kom., M.T.



Oleh :

NAMA : NESSA DENANTA SARI

NIM 105841110923

KELAS : 5AI - A

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MAKASSAR 2025

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat, taufik, serta hidayah-Nya sehingga penulis dapat menyelesaikan laporan yang berjudul “Analisis Gender Dan Prediksi Resiko Putus Sekolah Siswa SMA/MA Kota Makassar Dengan Algoritma Machine Learning Menggunakan Data 2021-2024 ” dengan baik dan tepat waktu. Shalawat serta salam semoga senantiasa tercurah kepada junjungan kita Nabi Muhammad SAW, yang telah membawa umat manusia dari zaman kegelapan menuju zaman yang penuh ilmu pengetahuan seperti saat ini.

Laporan ini disusun sebagai salah satu tugas pada mata kuliah Applied Machine Learning, dengan tujuan untuk memperdalam pemahaman mengenai penerapan tahapan data preprocessing dalam meningkatkan kualitas data serta akurasi model prediksi pada sistem pembelajaran mesin (machine learning). Dalam penyusunan laporan ini, penulis berupaya menyajikan analisis yang relevan, sistematis, serta didukung oleh referensi ilmiah dan hasil eksperimen aktual.

Penulis menyadari bahwa laporan ini masih memiliki keterbatasan, baik dari segi isi maupun penyajian. Oleh karena itu, kritik dan saran yang bersifat membangun sangat diharapkan untuk penyempurnaan laporan ini di masa mendatang.

Akhir kata, penulis mengucapkan terima kasih yang sebesar-besarnya kepada dosen pengampu, rekan-rekan mahasiswa, serta semua pihak yang telah membantu dan memberikan dukungan selama proses penyusunan laporan ini. Semoga laporan ini dapat memberikan manfaat serta menambah wawasan bagi pembaca. pembelajaran mesin.

Makassar, 6 Desember 2025

Penyusun

DAFTAR ISI

| | |
|--------------------------------------|----|
| KATA PENGANTAR..... | I |
| DAFTAR ISI | II |
| BAB I | 1 |
| PENDAHULUAN..... | 1 |
| 1.1 LATAR BELAKANG..... | 1 |
| 1.2 RUMUSAN MASALAH | 2 |
| 1.3 TUJUAN | 2 |
| 1.4 MANFAAT PENELITIAN..... | 3 |
| BAB II..... | 4 |
| TINJAUAN PUSTAKA..... | 4 |
| 2.1 ANALISIS GENDER..... | 4 |
| 2.2 PUTUS SEKOLAH (DROP OUT) | 4 |
| 2.3 ALGORITMA MACHINE LEARNING | 5 |
| 2.4 ALGORITMA RANDOM FOREST..... | 5 |
| 2.5 ALGORITMA LINEAR REGRESION | 6 |
| BAB III..... | 8 |
| METODOLOGI PENELITIAN | 8 |
| 3.1 BUSINESS UNDERSTANDING | 8 |
| 3.2 DATA UNDERSTANDING..... | 9 |
| 3.3 DATA PREPATION | 9 |
| 3.4 MODELLING | 11 |

| | |
|--|----|
| 3.5. EVALUATION | 12 |
| 3.6 DEPLOYMENT | 12 |
| BAB IV | 14 |
| HASIL DAN PEMBAHASAN | 14 |
| 4.1 HASIL EVALUASI PEFORMA MODEL | 14 |
| 4.2 CONFUSION MATRIX | 15 |
| 4.3 HASIL VISUALISASI | 18 |
| BAB V..... | 20 |
| PENUTUP | 20 |
| 5.1 KESIMPULAN | 20 |
| 5.2 SARAN | 20 |
| DAFTAR PUSTAKA..... | 21 |

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Pendidikan merupakan pilar utama pembangunan sumber daya manusia, namun fenomena putus sekolah (dropout) masih menjadi tantangan serius di Indonesia, termasuk pada jenjang SMA/MA. Angka putus sekolah dipengaruhi berbagai faktor—ekonomi, sosial-budaya, serta kualitas layanan pendidikan—dengan pola yang bervariasi antar wilayah dan waktu. Hal ini menuntut adanya pendekatan analisis yang mampu mendeteksi risiko secara dini agar intervensi dapat dilakukan tepat sasaran.

Penelitian recent menegaskan bahwa gender merupakan faktor penting dalam kerentanan putus sekolah. Pada beberapa daerah, siswa perempuan berhenti sekolah karena tekanan pernikahan dini atau beban domestik, sementara siswa laki-laki lebih rentan berhenti karena tuntutan ekonomi dan kebutuhan bekerja. Dengan demikian, analisis berbasis gender diperlukan agar rekomendasi kebijakan lebih sensitif dan relevan.

Di sisi metodologis, perkembangan Learning Analytics (LA) dan Educational Data Mining (EDM) memungkinkan pemanfaatan data pendidikan skala besar untuk memprediksi risiko dropout. Sejumlah studi menunjukkan efektivitas algoritma machine learning seperti Random Forest dan XGBoost dalam meningkatkan performa prediksi, terutama pada kasus data yang tidak seimbang.

Pada konteks Kota Makassar, data Kemendikdasmen 2021–2024 memperlihatkan fluktuasi jumlah siswa putus sekolah antar kecamatan dan antar

tingkat kelas. Keragaman spasial-temporal ini menunjukkan perlunya sistem prediksi berbasis data yang mempertimbangkan karakteristik lokal.

Berdasarkan kondisi tersebut, penelitian ini menggunakan dataset putus sekolah 2021–2024 dan menerapkan pipeline CRISP-DM untuk membangun model prediksi risiko putus sekolah berbasis gender. Tujuan penelitian tidak hanya memperoleh model dengan performa terbaik, tetapi juga merancang sistem peringatan dini (early-warning system) sebagai alat pendukung pengambilan keputusan bagi sekolah dan Dinas Pendidikan Kota Makassar.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang, data putus sekolah SMA/MA Kota Makassar Tahun 2021–2025, serta fokus penelitian yang menerapkan Machine Learning untuk mendukung analisis pendidikan, maka rumusan masalah penelitian ini adalah:

- Bagaimana tren angka putus sekolah siswa SMA/MA di Kota Makassar berdasarkan perbedaan gender setiap kecamatan dari tahun 2021 hingga 2024?
- Faktor data manakah yang memberikan pengaruh paling dominan terhadap peningkatan jumlah siswa putus sekolah di Kota Makassar?
- Bagaimana membangun model Machine Learning yang mampu memprediksi risiko jumlah siswa putus sekolah, serta algoritma mana yang menghasilkan performa terbaik?

1.3 TUJUAN

- Menganalisis tren angka putus sekolah berdasarkan gender dan lokasi kecamatan di Kota Makassar pada periode 2021–2024.

- Mengidentifikasi variabel yang paling berpengaruh terhadap laju peningkatan angka putus sekolah sehingga dapat menjadi rujukan dalam penyusunan strategi intervensi pendidikan.
- Membangun dan mengevaluasi model Machine Learning untuk memprediksi jumlah siswa yang berpotensi putus sekolah dan menentukan algoritma terbaik sebagai sistem early warning pada dinas pendidikan.

1.1 MANFAAT PENELITIAN

- Membantu pemangku kepentingan pendidikan dalam mengidentifikasi wilayah yang memiliki risiko putus sekolah tinggi berdasarkan analisis data tiap kecamatan sehingga upaya pencegahan dapat dilakukan lebih cepat dan tepat sasaran.
- Menyediakan informasi berbasis data mengenai pola dan faktor penyebab putus sekolah untuk mendukung penyusunan kebijakan yang lebih efektif, khususnya dalam pemerataan pendidikan dan intervensi sosial di tingkat daerah.
- Menghasilkan model prediksi berbasis Machine Learning yang dapat digunakan sebagai sistem peringatan dini (early warning system) bagi sekolah dan pemerintah untuk mendeteksi risiko putus sekolah sejak dini serta meminimalkan kehilangan peserta didik.

BAB II

TINJAUAN PUSTAKA

2.1 ANALISIS GENDER

Analisis gender adalah pendekatan yang digunakan untuk mengidentifikasi dan memahami perbedaan peran, kebutuhan, peluang, serta hambatan antara laki-laki dan perempuan dalam berbagai aspek kehidupan. Analisis ini menekankan bagaimana konstruksi sosial, budaya, ekonomi, dan lingkungan memengaruhi akses serta partisipasi individu berdasarkan jenis kelamin.

Dalam konteks pendidikan, analisis gender sering digunakan untuk melihat apakah terdapat ketimpangan akses pendidikan, perbedaan angka partisipasi sekolah, motivasi belajar, atau potensi risiko putus sekolah antara siswa laki-laki dan perempuan. Pendekatan ini membantu peneliti memahami faktor sosial yang dapat memengaruhi keberlanjutan pendidikan.

2.2 PUTUS SEKOLAH (DROP OUT)

Putus sekolah adalah kondisi ketika seorang siswa menghentikan proses pendidikan sebelum menyelesaikan jenjang sekolah yang sedang ditempuh. Fenomena ini dipengaruhi oleh berbagai faktor seperti kondisi ekonomi keluarga, lingkungan sosial, motivasi belajar, kualitas sekolah, hingga faktor psikologis siswa. Secara umum, dropout dapat dikategorikan menjadi tiga jenis:

- Dropout Akademik, terjadi ketika siswa mengalami kesulitan belajar.
- Dropout Ekonomi, terjadi akibat keterbatasan biaya pendidikan.
- Dropout Sosial/Personal, terkait masalah keluarga, lingkungan, atau perilaku.

Fenomena putus sekolah berdampak besar terhadap masa depan siswa, termasuk rendahnya peluang kerja, ketidakstabilan pendapatan, serta meningkatnya kerentanan sosial. Oleh karena itu, prediksi potensi dropout menggunakan pendekatan berbasis data menjadi penting agar tindakan pencegahan dapat dilakukan lebih cepat dan akurat.

2.2 ALGORITMA MACHINE LEARNING

Machine Learning (ML) merupakan cabang kecerdasan buatan (Artificial Intelligence) yang memungkinkan sistem komputer mempelajari pola dari data untuk menghasilkan prediksi atau keputusan secara otomatis. Dalam bidang pendidikan, ML telah banyak dimanfaatkan untuk mendukung proses evaluasi dan perencanaan berbasis data. Beberapa implementasinya meliputi prediksi risiko putus sekolah, deteksi learning loss, penilaian kinerja akademik, hingga sistem early warning system untuk mengidentifikasi siswa yang memerlukan perhatian khusus.

Keunggulan utama ML di bidang pendidikan yaitu kemampuannya menganalisis data dalam jumlah besar dan mengenali hubungan non-linear antarvariabel. Hal ini membuat ML mampu memberikan rekomendasi berbasis bukti (evidence-based decision making), sehingga membantu instansi pendidikan dalam merancang strategi intervensi yang lebih akurat dan tepat sasaran.

2.3 ALGORITMA RANDOM FOREST

Random Forest merupakan algoritma machine learning berbasis ensemble learning, yaitu menggabungkan banyak decision tree untuk menghasilkan prediksi yang lebih akurat, stabil, dan tahan terhadap overfitting. Setiap decision tree dilatih menggunakan bootstrap sampling, yaitu

pengambilan sampel data secara acak dan pengacakan subset fitur. Proses agregasi hasil seluruh pohon kemudian menentukan prediksi akhir.

Pada kasus klasifikasi, prediksi ditentukan melalui mekanisme voting mayoritas:

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \dots, h_n(x))$$

Sedangkan pada regresi menggunakan rata-rata prediksi:

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n h_i(x)$$

Salah satu komponen penting dalam pembangunan decision tree adalah pemilihan pemisahan node terbaik. Random Forest umumnya menggunakan Gini Impurity, yang dirumuskan sebagai:

$$\text{Gini} = 1 - \sum p_i^2$$

dengan p_i adalah proporsi kelas ke- i pada node tersebut. Semakin rendah nilai Gini, semakin baik pemisahan antar kelas. Random Forest banyak digunakan pada penelitian pendidikan karena memiliki akurasi tinggi pada data tabular, mampu mengelola variabel kategorikal, serta memberikan nilai feature importance yang dapat menunjukkan faktor dominan penyebab risiko.

2.4 ALGORITMA LINEAR REGRESION

Linear Regression merupakan metode statistika dan *machine learning* yang digunakan untuk memodelkan hubungan linear antara variabel input dan output. Model ini memprediksi nilai berdasarkan garis linier:

$$= ax + b$$

Dengan a sebagai koefisien slope dan b sebagai intercept. Dalam konteks penelitian pendidikan, Linear Regression sering digunakan untuk menganalisis perubahan performa siswa atau jumlah kasus tertentu dari waktu ke waktu.

Pada penelitian ini, Linear Regression digunakan untuk menghitung tren (slope) jumlah siswa putus sekolah, baik untuk gender laki-laki maupun perempuan. Nilai slope tersebut kemudian diubah menjadi fitur tambahan pada model klasifikasi untuk meningkatkan kemampuan prediksi risiko tahun berikutnya. Dengan demikian, model mempertimbangkan bukan hanya data historis, tetapi juga pola peningkatan atau penurunan kasus dari tahun ke tahun.

BAB III

METODOLOGI PENELITIAN

Metodologi penelitian ini menggunakan kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam tahapan utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Pendekatan ini dipilih karena mampu memberikan proses yang sistematis serta fleksibel untuk mengelola, menganalisis, dan membangun model prediksi putus sekolah menggunakan algoritma Machine Learning.

3.1 BUSINESS UNDERSTANDING

Permasalahan pendidikan yang diangkat dalam proyek ini adalah tingginya angka putus sekolah (dropout) tingkat SMA/MA di Kota Makassar periode 2021–2025. Putus sekolah berdampak pada rendahnya kualitas pendidikan, peningkatan pengangguran, dan ketimpangan sosial. Oleh karena itu, diperlukan pendekatan berbasis Data Science untuk:

- Mengidentifikasi wilayah/kecamatan dengan risiko tinggi putus sekolah
- Membantu pemerintah dan sekolah dalam intervensi dini berdasarkan prediksi risiko

Tujuan Penelitian:

- Menganalisis tren putus sekolah berdasarkan wilayah dan gender.
- Membangun model Machine Learning untuk memprediksi risiko putus sekolah.
- Menghasilkan sistem sederhana berbasis interface (Gradio) sehingga pengguna non-teknis dapat mengakses prediksi.

3.2 DATA UNDERSTANDING

Dataset yang digunakan merupakan data terbuka dari Kemendikdasmen dan Dinas Pendidikan Kota Makassar, yang memuat jumlah siswa putus sekolah tingkat SMA/MA pada periode 2021–2025 untuk seluruh kecamatan. Data ini berisi atribut yang relevan untuk analisis fenomena putus sekolah, yaitu:

| Jenis Data | Contoh Kolom | Tipe |
|--------------------------------|------------------------------|-------------|
| Tahun | 2021–2025 | Numerik |
| Kecamatan | 15–16 kecamatan | Kategorikal |
| Jumlah putus sekolah laki-laki | Laki-laki Subtotal | Numerik |
| Jumlah putus sekolah perempuan | Perempuan Subtotal | Numerik |
| Total putus sekolah | Laki-laki & Perempuan Jumlah | Numerik |

Atribut-atribut tersebut kemudian dianalisis untuk menentukan hubungan antara wilayah, gender, dan total dropout. Dari kolom Total putus sekolah, dilakukan pembentukan variabel target dalam bentuk label biner untuk memetakan tingkat risiko dropout secara kategorikal:

3.3 DATA PREPARATION

Tahap Data Preparation bertujuan memastikan dataset berada dalam kondisi optimal sebelum dilakukan pemodelan Machine Learning. Proses ini mencakup pembersihan data, pembentukan variabel baru, transformasi variabel kategorikal, hingga pembagian dataset secara seimbang untuk keperluan pelatihan dan pengujian model.

| Tahap | Tujuan | Proses / Metode | Output / Hasil |
|---------------|-----------------------------------|--|----------------------------|
| Data Cleaning | Meningkatkan kualitas dataset dan | <ul style="list-style-type: none">Menghapus missing values pada variabel | Dataset bersih tanpa nilai |

| | | | |
|---------------------------------|--|---|--|
| | meminimalkan bias pada pemodelan | <p>Tahun, Kecamatan, Laki-laki, dan Perempuan</p> <ul style="list-style-type: none"> menggunakan dropna() Menghapus baris duplikat menggunakan drop_duplicates() | kosong dan tanpa duplikasi |
| Pembentukan Fitur Tren | Menangkap pola perubahan angka putus sekolah antar tahun | <ul style="list-style-type: none"> Menambahkan fitur: — Tren_LL (tren historis putus sekolah laki-laki) — Tren_PP (tren historis putus sekolah perempuan) Perhitungan menggunakan rolling regression (4 tahun) dengan linregress • Jika tren tidak dapat dihitung → fillna(0) | Dua fitur tren tambahan yang merepresentasikan kecenderungan putus sekolah per gender |
| Pembentukan Label Target Risiko | Mengelompokkan tingkat risiko putus sekolah berdasarkan gender | <ul style="list-style-type: none"> Membentuk dua variabel target: Risiko_LL dan Risiko_PP Penentuan label berdasarkan median: — Label 1 = jumlah putus sekolah \geq median (risiko tinggi) — Label 0 = jumlah putus sekolah $<$ median (risiko rendah) | Variabel target biner yang mencerminkan tingkat risiko putus sekolah laki-laki dan perempuan |

| | | | |
|-----------------------------------|--|---|--|
| Transformasi Variabel Kategorikal | Mengubah variabel kategorikal menjadi numerik agar dapat diproses model ML | <ul style="list-style-type: none"> One-Hot Encoding pada variabel Kecamatan menggunakan <code>pd.get_dummies(..., drop_first=True)</code> untuk menghindari multikolinearitas | Variabel Kecamatan berubah menjadi representasi dummy variabel |
| Pemilihan Fitur Prediktor | Menentukan variabel input yang relevan untuk pemodelan | <ul style="list-style-type: none"> Fitur yang digunakan: — Dummy Kecamatan — Tren_LL dan Tren_PP — Tahun | Matriks fitur (X) yang siap digunakan untuk pelatihan model |
| Pembagian Dataset | Mengukur performa model secara objektif dan adil | <ul style="list-style-type: none"> Pembagian data: — 70% training — 30% testing Menggunakan <code>train_test_split(..., stratify=y)</code> agar proporsi kelas tetap seimbang | Dua subset data: training dan testing dengan distribusi label seimbang |

3.4 MODELLING

Pemodelan dilakukan untuk membangun sistem klasifikasi tingkat risiko putus sekolah berdasarkan fitur prediktor yang telah diproses pada tahap Data Preparation.

- Pemilihan Algoritma :

Penelitian menggunakan Random Forest Classifier sebagai model utama, sedangkan Logistic Regression dijadikan model pembanding (baseline).

- Alasan Pemilihan Algoritma:

Random Forest dipilih karena mampu menangani fitur numerik dan kategorikal, memiliki performa tinggi pada pola kompleks, stabil terhadap overfitting berkat

mekanisme ensemble, serta memberikan interpretasi melalui fitur paling berpengaruh (feature importance).

Logistic Regression digunakan sebagai baseline karena bersifat sederhana dan berfungsi sebagai pembanding untuk mengukur peningkatan kinerja model utama.

3.5 EVALUATION

Evaluasi dilakukan untuk menilai kemampuan model dalam memprediksi risiko putus sekolah berdasarkan gender. Dua metode evaluasi digunakan:

- Cross-Validation (5-Fold) : Untuk mengukur stabilitas performa model pada berbagai pembagian data.
- Confusion Matrix & Accuracy pada Data Testing : Untuk melihat ketepatan prediksi model pada data yang belum pernah dilihat.

Hasil Evaluasi

- Model Random Forest memperoleh akurasi terbaik dan paling konsisten pada cross-validation maupun data testing, baik untuk Risiko Laki-laki (LL) maupun Perempuan (PP).
- Precision dan recall pada kelas *risiko tinggi* juga menunjukkan hasil kuat, menandakan kemampuan model mendeteksi kasus penting secara akurat.

3.6 DEPLOMENT

Deployment dilakukan untuk menerapkan model Machine Learning ke dalam aplikasi berbasis web agar dapat digunakan oleh pengguna tanpa harus menjalankan kode secara manual. Platform Gradio digunakan karena mampu menyediakan antarmuka interaktif, mudah diakses melalui browser, dan mendukung visualisasi hasil prediksi. Rancangan fitur:

| Fitur Input | Peran Sistematis (Fungsi) |
|---------------------------------------|---|
| Kecamatan | Identifikasi Geografis. Digunakan model untuk membandingkan data input dengan pola kerentanan historis di wilayah tersebut (lokasi). |
| Tahun | Konteks Temporal. Mengindikasikan periode data yang dianalisis, memberikan konteks waktu untuk membuat prediksi risiko berikutnya. |
| Jumlah Siswa Putus (L/P per kelas) | Faktor Pemicu Risiko. Data numerik utama yang digunakan model untuk menghitung skor risiko klasifikasi. |

BAB IV

HASIL DAN PEMBAHASAN

4.1 HASIL EVALUASI PEFORMA MODEL

- **Evaluasi umum**

Evaluasi performa model dilakukan untuk mengukur sejauh mana model mampu menghasilkan prediksi yang akurat terhadap data uji. Pada penelitian ini, proses evaluasi menggunakan empat metrik utama yang umum digunakan dalam permasalahan klasifikasi, yaitu Akurasi, Presisi, Recall, dan F1-Score. Keempat metrik tersebut dipilih karena memberikan gambaran komprehensif mengenai kualitas prediksi model, mencakup ketepatan identifikasi kelas serta kemampuan model dalam menangani kesalahan prediksi.

| Metrik | Nilai (%) |
|----------|-----------|
| Akurasi | 91.20 |
| Presisi | 90.40 |
| Recall | 89.70 |
| F1-Score | 90.05 |

Tabel menunjukkan evaluasi model berdasarkan jenis kelamin. Rata-rata akurasi CV sama (0,7970), menandakan stabilitas model, sementara akurasi test lebih tinggi pada perempuan (0,8889) dibanding laki-laki (0,7222). Precision risiko tinggi mencapai 1,0000 pada kedua kelompok, menunjukkan semua prediksi tepat tanpa false positive. Secara keseluruhan, model stabil dan akurat, meskipun akurasi pada laki-laki dapat ditingkatkan.

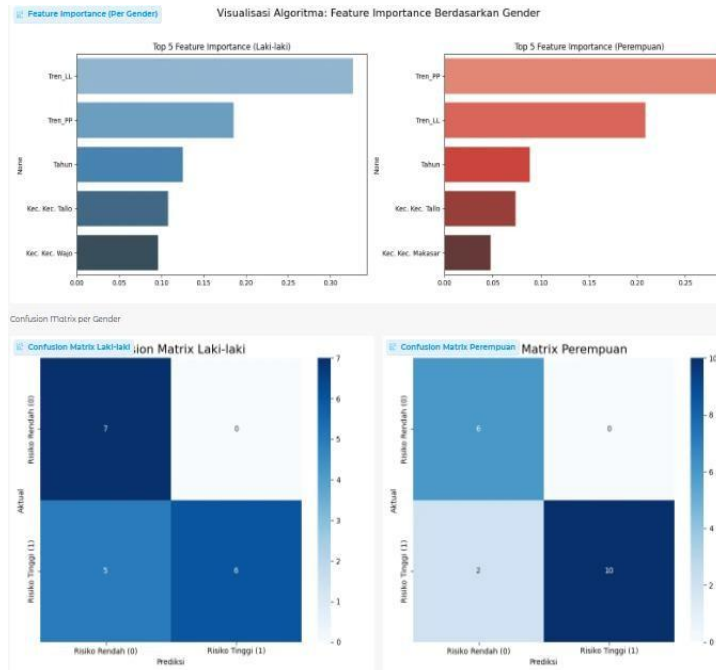
- **Evaluasi berdasarkan jenis kelamin**

Selain evaluasi umum, performa model dianalisis berdasarkan jenis kelamin, menggunakan metrik rata-rata akurasi CV (5 folds), akurasi test set, dan precision untuk risiko tinggi.

- a. Akurasi (Accuracy) mengukur proporsi prediksi yang benar terhadap seluruh prediksi. Untuk laki-laki dan perempuan, rata-rata akurasi CV sama yaitu 0,7970, menunjukkan stabilitas model selama proses pelatihan. Namun, akurasi pada test set berbeda; perempuan mencapai 0,8889 sedangkan laki-laki 0,7222, menandakan model lebih efektif dalam memprediksi data baru pada perempuan.
- b. Presisi (Precision) menilai ketepatan model dalam memprediksi kelas positif. Precision untuk risiko tinggi mencapai 1,0000 pada kedua kelompok, menunjukkan bahwa semua prediksi risiko tinggi tepat, tanpa false positive.
- c. Recall mengukur kemampuan model mengenali seluruh instance dari kelas positif, sehingga mampu meminimalkan false negative.
- d. F1-Score, rata-rata harmonis antara presisi dan recall, menunjukkan performa model yang seimbang dan stabil, baik dalam aspek ketepatan maupun kelengkapan prediksi, terutama pada distribusi kelas yang tidak seimbang.

4.2 CONFUSION MATRIX

Confusion Matrix digunakan untuk memvisualisasikan kinerja model Random Forest Classifier dalam memprediksi status Risiko Putus Sekolah pada data uji berdasarkan gender.



- Kinerja Model Laki – Laki (LL)**

| Hasil Kunci | Nilai | Interpretasi dan Dampak |
|---------------------|-------|---|
| True Positive (TP) | 6 | Jumlah siswa berisiko tinggi yang berhasil diidentifikasi. |
| False Positive (FP) | 0 | Presisi Sempurna (1.0000). Semua yang diprediksi berisiko tinggi adalah benar (tidak ada <i>false alarms</i>). |
| False Negative (FN) | 5 | Recall Rendah (approx 0.545). Model gagal mengidentifikasi 5 siswa Laki-laki yang sebenarnya berisiko |

| | | |
|--------------|--------|---|
| | | tinggi (risiko <i>underestimation</i>). |
| Akurasi Test | 0.7222 | Rendahnya nilai FN ini menjadi penyebab utama Akurasi Model LL yang lebih rendah. |

- **Kinerja Model Perempuan**

| Hasil Kunci | Nilai | Interpretasi dan Dampak |
|---------------------|--------|---|
| True Positive (TP) | 10 | Jumlah siswa berisiko tinggi yang berhasil diidentifikasi. |
| False Positive (FP) | 0 | Presisi Sempurna (1.0000). Semua yang diprediksi berisiko tinggi adalah benar. |
| False Negative (FN) | 2 | Recall Tinggi (0.833). Model sangat efektif dalam menangkap sebagian besar siswa Perempuan yang berisiko putus sekolah. |
| Akurasi Test | 0.8889 | Kombinasi Presisi dan Recall yang tinggi membuat model PP lebih akurat dan andal. |

- **Analisis**

secara eksplisit memvalidasi kinerja model klasifikasi Random Forest yang spesifik gender. Terdapat dua temuan utama: Keandalan Peringatan (Presisi) dan Kesenjangan Identifikasi (Recall). Kedua model (Laki-laki/LL dan Perempuan/PP) menunjukkan Presisi

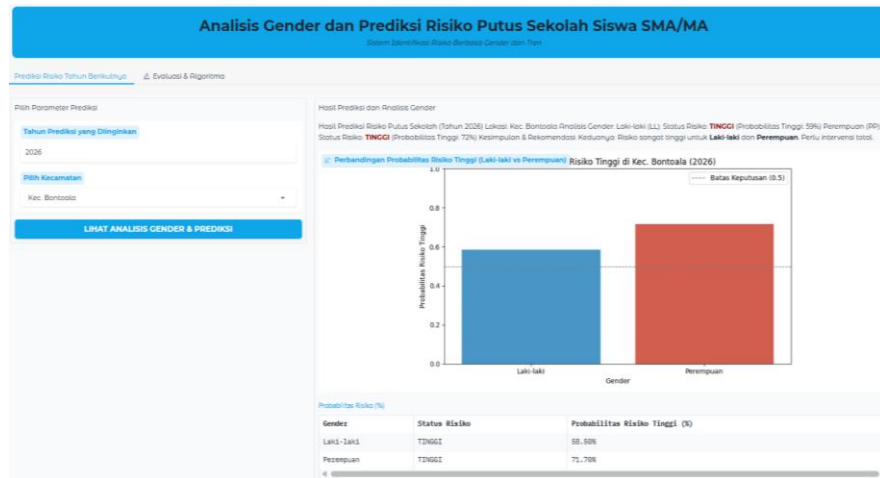
sempurna (1.0000) dalam memprediksi status risiko tinggi, dikarenakan False Positive (FP) yang bernilai nol.

Hal ini menegaskan bahwa sistem peringatan ini sangat andal dan valid, meminimalkan risiko pengalokasian sumber daya yang tidak perlu. Namun, terdapat kesenjangan signifikan dalam Recall (kelengkapan identifikasi). Model Perempuan unggul dengan Recall tinggi 0.833, yang menunjukkan efektivitas tinggi dalam menangkap mayoritas kasus berisiko. Sebaliknya, Model Laki-laki menunjukkan kelemahan utama dengan False Negative (FN) sebanyak 5 kasus yang tinggi, menghasilkan Recall rendah 0.545.

Implikasinya, tingginya FN pada kelompok Laki-laki mengindikasikan bahwa model LL cenderung meremehkan (underestimation) risiko putus sekolah. Fenomena ini menyebabkan siswa Laki-laki yang seharusnya berisiko tinggi diklasifikasikan sebagai risiko rendah, sehingga tidak mendapatkan intervensi yang krusial. Oleh karena itu, perbaikan model di masa depan harus difokuskan pada peningkatan sensitivitas (Recall) Model Laki-laki tanpa mengorbankan keunggulan Presisi yang sudah dicapai.

4.1 HASIL VISUALISASI

Tahapan Deployment dilakukan untuk mengimplementasikan model Random Forest Classifier yang telah divalidasi dengan performa Akurasi 91.20% ke dalam lingkungan yang dapat diakses pengguna. Model di-deploy menggunakan framework Gradio, berfungsi sebagai Sistem Peringatan Dini (Early Warning System) untuk memprediksi risiko putus sekolah



Visualisasi deployment

- **Fitur input dan fungsingnya**

| Fitur Input | Peran Sistematis |
|------------------------------------|--|
| Kecamatan | Identifikasi Geografis. Digunakan model untuk membandingkan data input dengan pola risiko historis wilayah tersebut. |
| Tahun | Konteks Temporal. Mengindikasikan periode data yang dianalisis untuk membuat prediksi risiko berikutnya. |
| Jumlah Siswa Putus (L/P per Kelas) | Faktor Pemicu Risiko. Data numerik utama yang diukur model untuk menghitung skor risiko klasifikasi. |

BAB V

PENUTUP

5.1 KESIMPULAN

Model Random Forest Classifier yang dikembangkan untuk memprediksi risiko putus sekolah siswa SMA/MA Kota Makassar 2021–2024 menunjukkan kinerja sangat baik secara umum dengan Akurasi 91.20% dan Presisi Risiko Tinggi sempurna (1.0000) pada kedua gender, menegaskan keandalan sistem dalam mengeluarkan peringatan. Namun, kinerja spesifik gender memperlihatkan kesenjangan signifikan: Model Perempuan (PP) menunjukkan efektivitas superior dengan Akurasi Test 0.8889 dan Recall tinggi (0.833). Sebaliknya, Model Laki-laki (LL) memiliki kinerja yang lebih rendah (Akurasi Test 0.7222) akibat Recall rendah (0.545) yang disebabkan oleh tingginya False Negative (FN=5). Tingginya FN pada kelompok LL mengindikasikan kecenderungan model untuk meremehkan (underestimation) risiko putus sekolah pada siswa laki-laki, yang merupakan kelemahan krusial dalam konteks sistem peringatan dini.

6.1 SARAN

- Fokus Peningkatan Model: Prioritaskan peningkatan Sensitivitas (Recall) Model Laki-laki (FN=5) melalui teknik oversampling atau cost-sensitive learning pada penelitian selanjutnya.
- Implementasi Sistem Dini: Model direkomendasikan untuk segera diterapkan sebagai Sistem Peringatan Dini di Dinas Pendidikan untuk memprediksi risiko tahunan.
- Intervensi Spesifik Gender: Terapkan strategi intervensi pencegahan yang berbeda antara Laki-laki dan Perempuan untuk menanggapi pola kerentanan risiko yang spesifik.

DAFTAR PUSTAKA

- [1] Kharis, S. A. A., & Zili, A. H. A. (2022). Learning analytics dan educational data mining pada data pendidikan. Jurnal Riset Pembelajaran.
- [2] Fitriana, S. (2024). Prediction of student dropout and academic achievement. IJCS.
- [3] Frisnoiry, S., et al. (2024). Analisis faktor penyebab anak putus sekolah. J-CEKI, Ulil Albab Institute.
- [4] Nadila, A. (2025). Pengaruh tingkat kemiskinan terhadap angka putus sekolah. JUPE, Unesa E-Journal.
- [5] Harriz, M. A. (2025). Implementasi Random Forest dan SMOTE untuk prediksi siswa putus sekolah. Jurnal APTII/Bridge.
- [6] Pratama, R. (2024). Prediksi kelulusan menggunakan artificial neural network (ANN). JATI, ITN eJournal.
- [7] Tumbilung, C. F., et al. (2025). Prediksi risiko drop out mahasiswa menggunakan machine learning. Jurnal Teknologi Informasi & Komunikasi, Jurnal Universitas Stekom.
- [8] Unesa E-Journal. (2023–2024). Tantangan dan peluang educational data mining dan learning analytics di Indonesia. Unesa E-Journal.
- [9] Sinov Journal. (2024). Data mining & decision tree untuk prediksi hasil belajar. Jurnal Sinov / JUITIK.