

BAB II

TINJAUAN PUSTAKA

2.1. Landasan Teori

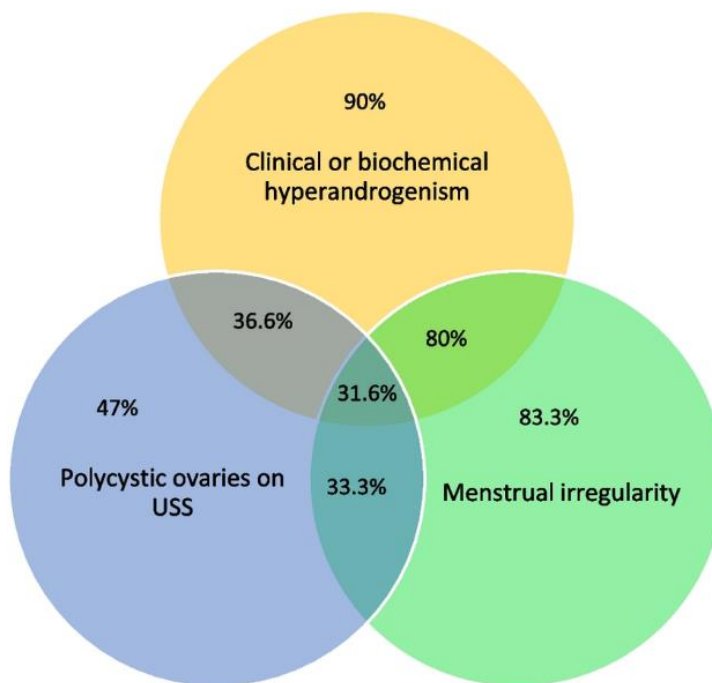
2.1.1 *Polycystic Ovary Syndrome (PCOS)*

Polycystic Ovary Syndrome (PCOS) merupakan salah satu gangguan metabolik perempuan yang paling umum yang mempengaruhi perempuan usia subur, dengan prevalensi global berkisar antara 5% hingga 20%, tergantung pada populasi dan kriteria diagnosis yang digunakan (Di Michele dkk., 2025; Johnson dkk., 2025; Shekarian dkk., 2025).

Saat ini belum diketahui secara pasti penyebab PCOS. Biasanya terjadi karena pola hidup yang buruk dan tidak teratur. Perempuan baru akan menyadari terkena PCOS ketika akan memiliki keturunan. Padahal, salah satu cirinya, sangat mudah dikenali, misalnya tidak lancarnya siklus menstruasi. Gejala yang sering ditemukan pada pasien *Polycystic Ovary Syndrome* (PCOS) di antaranya siklus menstruasi, tumbuh rambut berlebih (hirsutisme), jerawat yang sulit diatasi, penambahan berat badan, kesulitan untuk hamil, serta adanya kista kecil di ovarium (Sadeghi dkk., 2022; Singh dkk., 2023). Namun, gejala ini dapat berbeda-beda pada tiap individu, sehingga diagnosis seringkali memerlukan evaluasi medis secara menyeluruh.

Diagnosis *Polycystic Ovary Syndrome* (PCOS) umumnya merujuk pada Kriteria Rotterdam yang diperkenalkan pada tahun 2003 oleh ESHRE (*European Society of Human Reproduction and Embryology*) dan ASRM

(*American Society for Reproductive Medicine*) (Baba, 2025; Ghafari dkk., 2025). Gambar 2.1. menampilkan Kriteria Rotterdam.



Gambar 2. 1 Kriteria Rotterdam

Kriteria ini menetapkan bahwa seseorang didiagnosis PCOS apabila memenuhi setidaknya dua dari tiga kondisi berikut:

1. Oligo atau Anovulasi

Oligo/anovulasi ialah kondisi ketika ovulasi terjadi secara tidak teratur, atau bahkan tidak terjadi sama sekali. Siklus menstruasi yang jarang (oligomenorea) atau tidak terjadi sama sekali (amenorea) adalah indikasi terjadinya gangguan ini. Umumnya, kondisi ini ditandai dengan siklus menstruasi yang lebih dari 35 hari atau terjadi kurang dari 8 kali dalam setahun (Ghafari dkk., 2025; Gorzko dkk., 2025). Infertilitas pada

Perempuan dengan PCOS biasanya disebabkan oleh ketidakaturan ovulasi.

2. Hiperandrogenisme

Hiperandrogenisme dapat ditunjukkan secara klinis dan biokimia. Secara klinis, gejala yang ditunjukkan adalah pertumbuhan seperti pertumbuhan rambut (*hirsutisme*), jerawat yang parah dan menetap, serta kerontokan rambut di kepala (*alopecia*) (Baba, 2025).

Sedangkan secara biokimia, melalui pemeriksaan laboratorium yang menunjukan kadar hormon androgen yang tinggi, seperti testosteron, DHEAS, atau androstenedione (Ghafari dkk., 2025).

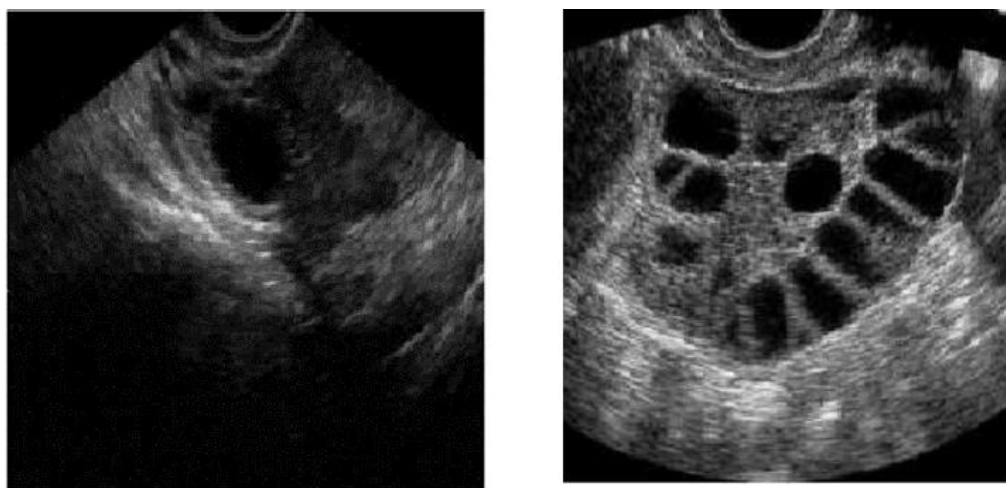
Tidak semua pasien menunjukkan tanda-tanda klinis yang nyata. Sehingga, pemeriksaan laboratorium tetap diperlukan untuk mengonfirmasi hiperandrogenisme biokimia (Gorzko dkk., 2025).

3. Morfologi Ovarium Polikistik (PCOM)

Polycystic Ovarian Morphology (PCOM) ditentukan berdasarkan pemeriksaan ultrasonografi (USG), terutama dengan *transvaginal ultrasound* yang menggunakan *probe* frekuensi tinggi (≥ 8 MHz) (Al-Hamzawi, 2025). Kriteria morfologi ovarium polikistik mencakup:

1. Jumlah folikel kecil ≥ 20 dengan ukuran 2-9mm pada salah satu atau kedua ovarium, atau
2. Volume ovarium lebih dari 10mL (Ghafari dkk., 2025).

Pendeteksian karakteristik ovarium polikistik, tidak hanya berdasarkan pemeriksaan USG, melainkan kadar *Anti-Mullerian Hormone* (AMH) yang tinggi juga dapat digunakan sebagai indikator alternatif untuk mendeteksi karakteristik ovarium polikistik, terutama pada kondisi pemeriksaan USG sulit dilakukan (Gorzko dkk., 2025). Gambar 2.2. menunjukkan kondisi rahim yang normal dan rahim pengidap PCOS.



Gambar 2. 2 Rahim Normal dan Rahim PCOS

Menurut beberapa studi, perempuan baru akan menyadari dirinya terkena PCOS setelah menghadapi kesulitan untuk mendapatkan keturunan. *Polycystic Ovary Syndrome* (PCOS) merupakan salah satu faktor utama infertilitas atau anovulasi pada perempuan. Oligo atau anovulasi merupakan hambatan utama pada kehamilan.

Polycystic Ovary Syndrome (PCOS) tidak hanya berdampak signifikan terhadap kesehatan reproduksi perempuan. Penyakit ini juga erat

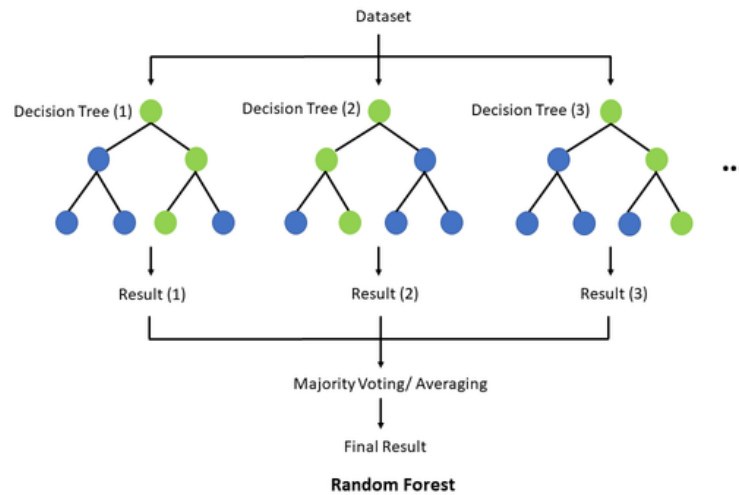
kaitannya dengan penyakit kardiovaskular (Amiri dkk., 2025; Xu & Qiao, 2022). Penyakit ini berhubungan dengan dislipidemia, resistensi insulin, peradangan sistemik, dan disfungsi endotel (Henney dkk., 2024; Su dkk., 2025). Bahkan untuk Perempuan dengan PCOS yang tidak obesitas, berisiko terkena penyakit jantung koroner dan infark miokard (Gersh dkk., 2025).

Gejala-gejala *Polycystic Ovary Syndrome* (PCOS) seperti jerawat parah, hirsutisme, kegemukan, dan infertilitas juga dapat berdampak signifikan pada kesehatan mental, seperti depresi, gangguan kecemasan, gangguan citra tubuh, hingga gangguan makan (Geller dkk., 2025; Hofmann dkk., 2025). Tidak sedikit, perempuan dengan PCOS yang mengalami stres berat dan merasa kehilangan kontrol terhadap kondisi tubuhnya. Kombinasi beban fisik dan psikologis ini menyebabkan penurunan kualitas hidup yang signifikan (Dybciak dkk., 2022; Lee dkk., 2024; Yang dkk., 2025).

2.1.2 Random Forest

Random Forest merupakan salah satu algoritma ensambel learning berbasis pohon keputusan. Random Forest pertama kali diperkenalkan oleh Leo Breiman pada tahun 2001. Random Forest bekerja dengan cara membentuk banyak pohon keputusan secara acak, lalu hasil prediksi tiap dari tiap pohon keputusan (*decision tree*) digabungkan untuk menghasilkan keputusan akhir dengan cara voting (klasifikasi) atau rata-rata (regresi) (Schonlau & Zou, 2020). Sehingga menjadikan Random Forest bersifat non-parametrik, fleksibel, dan mampu menangani hubungan non-linear antar

variabel (Hatwell dkk., 2020; Langsetmo dkk., 2023). Arsitektur Random Forest ditampilkan pada Gambar 2.3.



Gambar 2. 3 Arsitektur Random Forest

Random Forest bekerja dengan cara membangun banyak *decision tree* (pohon keputusan) menggunakan *aggregating* atau *bagging*. Setiap pohon dilatih berdasarkan data latih yang diambil secara acak dengan pengembalian (bootstrap dan sampel), kemudian pada setiap pemisahan node, hanya sebagian fitur yang dipilih secara acak. Proses ini mengurangi korelasi antar pohon dan meningkatkan kemampuan generalisasi model. Pada klasifikasi, Random Forest menggunakan metrik *Gini Impurity* untuk menentukan pemisahan (*split*) terbaik pada suatu node. Gini Impurity, dihitung dengan Rumus (1).

$$Gini = 1 - \sum_{\{i=1\}}^{\{C\}} p_i^2 \quad (1)$$

Dengan penjelasan sebagai berikut:

C : Jumlah kelas

p_i : Jumlah proporsi data pada kelas ke- i pada node tersebut

Nilai Gini yang semakin kecil, semakin murni node tersebut, artinya semakin baik pemisahannya.

Prediksi akhir dilakukan apabila seluruh pohon keputusan telah terbentuk, kemudian dilakukan penggabungan hasil dari semua pohon. Pada kasus klasifikasi, voting mayoritas dilakukan dengan menggunakan Rumus (2).

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \dots, h_T(x)) \quad (2)$$

Sedangkan, pada kasus regresi, prediksi akan ditentukan berdasarkan rata-rata nilai yang dihasilkan oleh semua pohon. Dihitung dengan menggunakan Rumus (3).

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (3)$$

Jumlah total pohon direpresentasikan dengan T , sedangkan $h_t(x)$ merupakan prediksi dari pohon ke- t .

Untuk mengevaluasi performa model secara efisien tanpa perlu data validasi terpisah, Random Forest menggunakan metode *Out-of-Bag* sebagai metode estimasi akurasi oleh. Metode ini menggunakan data yang tidak

terpilih pada proses pemilihan pohon tertentu, dan dirumuskan pada Rumus (4).

$$OOB\ Error = \frac{1}{n} \sum_{i=1}^n I(\widehat{y}_1^{OOB} \neq y_i) \quad (4)$$

Simbol I menyatakan fungsi indikator yang bernilai 1 apabila prediksi tidak sesuai dengan label asli, dan 0 apabila sesuai. Nilai n mengacu pada jumlah total; data observasi.

2.1.3 *Extreme Gradient Boosting (XGBoost)*

Boosting merupakan teknik ensambel yang memiliki tujuan untuk menggabungkan beberapa model lemah (*weak learners*) secara iteratif guna membentuk model yang kuat. Gradient Boosting merupakan salah satu pendekatan *boosting* yang paling dikenal, pendekatan ini dikembangkan oleh Friedman pada tahun 2001. Gradient Boosting melatih model baru untuk memperbaiki kesalahan model sebelumnya dengan cara mengoptimasi fungsi loss (Bentéjac dkk., 2021).

Fungsi prediksi pada *gradient boosting* ditampilkan dengan Rumus (5).

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (5)$$

Dengan penjelasan sebagai berikut :

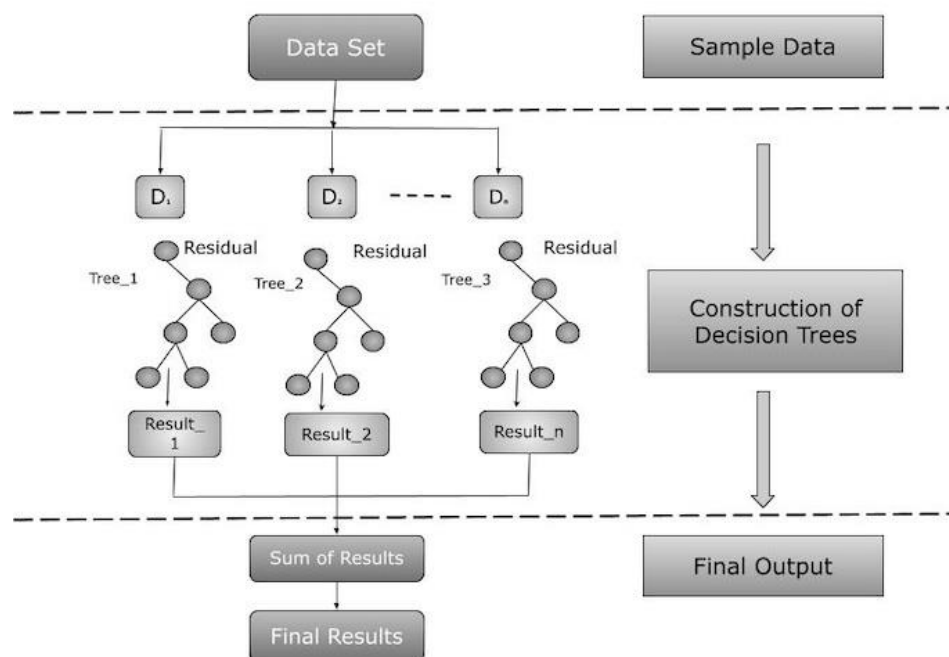
$h_m(x)$: Model lemah ke-m (biasanya pohon keputusan yang dangkal)

γ_m : Bobot kontribusi dari model ke-m

M : Jumlah iterasi atau model

Boosting memperbaiki kesalahan prediksi sebelumnya pada setiap iterasi dengan meminimalkan fungsi *loss*. Misalnya pada regresi, fungsi loss dapat berupa *mean squared error*, dan pada klasifikasi bisa berupa *log loss* (Sahin, 2020).

Salah satu pengembangan dari gradient boosting yang paling dikenal ialah XGBoost. Diperkenalkan oleh Chen dan Guestrin pada tahun 2016, dengan fokus pada efisiensi, skalabilitas, dan regularisasi (Bentéjac dkk., 2021). Gambar 2.4 menunjukkan arsitektur XGBoost.



Gambar 2. 4 Arsitektur XGBoost

XGBoost bekerja sesuai dengan prinsip boosting, yang bekerja dengan cara membangun model prediksi secara bertahap dengan cara menambahkan pohon keputusan satu per satu. Setiap pohon baru dilatih untuk memperbaiki kesalahan dari pohon-pohon sebelumnya.

Tahapan pertama, model akan memulai prediksi awal yang konstan. Biasanya, ini merupakan rata-rata dari target untuk regresi, atau log-odds untuk klasifikasi. Selanjutnya, XGBoost akan melakukan proses pembelajaran secara iteratif. Pada setiap iterasi t , dilakukan langkah sebagai berikut

1. Hitung gradien dan hessian

Pada setiap data *training*, nilai turunan pertama (*gradien*) dan turunan kedua (*hessian*) dihitung dari fungsi loss terhadap prediksi sebelumnya (P. Zhang dkk., 2022). Rumus perhitungan gradien ditunjukkan pada Rumus (6), sedangkan perhitungan hessian ditunjukkan pada Rumus (7).

$$g_i = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i} \quad (6)$$

$$h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^2} \quad (7)$$

2. Membangun pohon baru

Pohon keputusan $f_{t(x)}$ baru dibangun berdasarkan nilai gradien dan hessian. Pohon tersebut meminimalkan fungsi objektif. Pohon dibentuk dengan cara mencari nilai pembagian (split) terbaik yang memberikan *gain* maksimal terhadap fungsi objektif (Wiens dkk., 2025).

3. Memperbarui prediksi

Prediksi keseluruhan diperbarui dengan cara menambahkan *output* dari pohon baru, prediksi terbaru dihitung dengan menggunakan Rumus (8).

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t_1)} + \eta f_t(x_i) \quad (8)$$

XGBoost juga menambahkan regularisasi pada fungsi objektif, yang bertujuan untuk mencegah overfitting (P. Zhang dkk., 2022). Regularisasi ini memperhatikan jumlah daun (leaf) pada pohon, besar bobot pada tiap daun, serta kompleksitas struktur pohon. Regularisasi ditampilkan pada Rumus (9).

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (9)$$

Apabila setelah beberapa iterasi model pada data validasi tidak menunjukkan akurasi yang baik. Proses pelatihan dihentikan untuk mencegah *overfitting* (*early stopping*) (Wiens dkk., 2025).

2.1.4 *Synthetic Minority Over-sampling Technique (SMOTE)*

Synthetic Minority Over-sampling Technique (SMOTE) merupakan salah satu teknik *oversampling* yang digunakan untuk mengatasi masalah ketidakseimbangan kelas, atau kondisi ketika jumlah sampel pada kelas mayoritas jauh dengan kelas minoritas. Pendekatan ini dinilai jauh lebih efektif dibandingkan dengan duplikasi data minoritas secara langsung karena dapat menyebabkan *overfitting* (Hussein dkk., 2019).

SMOTE bekerja dengan cara menghasilkan sampel sintetis baru dari kelas minoritas. Proses ini dilakukan dengan cara interpolasi linear antara suatu data minoritas dengan salah satu dari k tetangga terdekatnya. Secara matematis, prosesnya dapat dirumuskan pada Rumus (10).

$$x_{\text{synthetic}} = x_i + \delta \cdot (x_{zi} - x_i) \quad (10)$$

Berikut penjelasan dari rumus SMOTE pada Rumus (10). x_i adalah *instance* dari kelas minoritas, sedangkan x_{zi} adalah salah satu dari k -tetangga terdekat, dan δ adalah bilangan acak antara 0 dan 1. Dengan formula ini sangat mungkin apabila terciptanya variasi data baru yang berada pada distribusi kelas minoritas yang sama, sehingga dapat memperluas ruang representasi fitur secara lebih realistis (Gholampour, 2024).

2.1.5 *Shapley Additive ExPlanations (SHAP)*

Shapley Additive Explanations (SHAP) merupakan salah satu pendekatan *explainable AI (XAI)* yang dibuat untuk menjelaskan *output machine learning* secara kuantitatif dengan cara mendistribusikan

kontribusi dari setiap fitur pada hasil prediksi. Pendekatan ini dikembangkan berdasarkan teori nilai *Shapley* dari teori permainan kooperatif oleh Lloyd Shapley (1953). SHAP memiliki konsistensi yang kuat serta distribusi kontribusi yang adil (Angelov dkk., 2021).

SHAP merupakan pendekatan berbasis metode *attribution feature additive* yang menggabungkan kelebihan dari berbagai metode interpretasi. Prediksi model $f(x)$ direpresentasikan pada Rumus (11).

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i \quad (11)$$

Dengan penjelasan sebagai berikut:

- ϕ_0 : *Expected value* dari model ketika semua fitur tidak disertakan (nilai dasar)
- ϕ_i : Kontribusi dari fitur ke- i terhadap prediksi
- M : Jumlah fitur

Rumus asli nilai *Shapley* untuk menghitung kontribusi fitur i direpresentasikan pada Rumus (12).

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (12)$$

Dengan penjelasan sebagai berikut:

- N : Himpunan semua fitur
- S : Subset dari N yang tidak mengandung i .

$f(S)$: Nilai prediksi model ketika hanya fitur dalam subset S digunakan

Rumus nilai *Shapley* menghitung kontribusi marjinal fitur ke- i terhadap seluruh subset yang mungkin dari fitur lainnya. Dengan mempertimbangkan semua kombinasi subset dan menghitung selisih antara hasil prediksi sebelum dan sesudah fitur i ditambahkan, nilai ini memberikan kontribusi yang adil dan konsisten terhadap output model (K. Zhang dkk., 2020).

Nilai SHAP merepresentasikan besarnya kontribusi setiap fitur terhadap suatu hasil prediksi spesifik. Nilai ini sifatnya lokal dan unik pada setiap observasi. Apabila nilai SHAP suatu fitur positif, artinya fitur tersebut berkontribusi meningkatkan probabilitas terhadap kelas target. Sedangkan, nilai negatif artinya pengurangan kontribusi pada kelas tersebut.

Srinivasu dkk., 2022 mengungkapkan bahwa SHAP memiliki tiga properti utama yang menjadi faktor interpretasi yang kuat. Pertama, jumlah nilai SHAP semua fitur ditambah nilai dasar sama dengan prediksi model (*local accuracy*). Kedua, fitur yang tidak disertakan memiliki kontribusi nol (*missingness*). Ketiga, peningkatan kontribusi fitur terhadap model akan diikuti peningkatan nilai SHAP (*consistency*).

Menurut Angelov dkk., 2021 SHAP mendukung dua pendekatan interpretasi, yakni pendekatan lokal dan global. Pendekatan interpretasi lokal ialah yang menjelaskan mengapa suatu prediksi dibuat pada individu. Biasanya divisualisasikan dengan *force plot* dan *waterfall plot*. Sedangkan,

pendekatan global ialah pendekatan yang menjelaskan keseluruhan perilaku model terhadap suatu data. Biasanya divisualisasikan dengan *summary plot* dan *bar plot*.

2.1.6 Preprocessing Data

Preprocessing data adalah sekumpulan proses yang dilakukan sebelum model dilatih, tahapan ini adalah tahapan yang krusial. Data mentah yang diperoleh dari sumber biasanya terdapat nilai kosong, fitur yang tidak diperlukan, kelas yang tidak seimbang, dan permasalahan data lainnya. Oleh karena itu, tahapan ini bertujuan untuk menjadikan data layak dan optimal untuk digunakan dalam proses pelatihan model. Data yang kotor, banyak nilai yang kosong, dan kelas yang tidak seimbang mengakibatkan prediksi yang dihasilkan oleh model bias, oleh sebab itu tahapan ini sangat menentukan kualitas model (Fan dkk., 2021). Proses preprocessing data pada penelitian ini meliputi :

1. Penghapusan Kolom yang Tidak Relevan

Proses pemilihan fitur (*feature selection*) adalah sebuah proses dalam preprocessing yang bertujuan untuk menyeleksi fitur-fitur yang informatif dan menghapus fitur-fitur yang tidak relevan, sehingga performa model meningkat dan model dapat bekerja dengan efisien (Pudjihartono dkk., 2022).

Dalam sebuah data, khususnya data medis, seringkali terdapat kolom-kolom yang tidak memiliki nilai prediktif terhadap variabel target. Biasanya terdapat kolom-kolom seperti ID Pasien, *timestamp*, atau data-

data lainnya yang tidak memiliki makna terhadap proses klasifikasi. Hal ini berpengaruh signifikan pada model, risikonya ialah overfitting dan *noise*.

2. *Missing Value Handling*

Data-data dengan nilai kosong sering ditemukan dalam sebuah dataset. Hal ini dapat menyebabkan gangguan pada proses pelatihan model *machine learning*. Dalam konteks komputasi, hal ini biasa disebut dengan *missing values*.

Missing values biasa diatasi dengan metode imputasi sederhana, yaitu dengan cara mengganti nilai yang hilang dengan menggunakan nilai statistik dari kolom tersebut, misalnya dengan mean, modus, dan median. Metode imputasi data dengan mean, modus, dan median adalah metode yang sering digunakan karena mudah diimplementasikan dan cukup efektif, terutama pada proporsi data yang hilang relatif kecil (Hunt, 2017). Namun, penelitian ini tidak menggunakan ketiga pendekatan tersebut, melainkan dengan cara menghapus baris data yang terdapat *missing value*, dengan catatan proporsi data yang hilang relatif kecil sehingga penghapusan tidak mengakibatkan kehilangan informasi yang signifikan. Selain itu, pendekatan ini dipilih untuk menjaga integritas data medis agar tetap merepresentasikan pasien.

3. *Imbalance Data Handling*

Dataset seringkali bermasalah pada keseimbangan kelas. Distribusi data yang tidak seimbang disebabkan karena kondisi pada suatu kelas lebih besar daripada kelas lainnya. Kondisi ini seringkali menyebabkan model

menjadi bias terhadap kelas mayoritas, dan cenderung mengabaikan performa klasifikasi pada kelas minoritas sehingga performa pada minoritas buruk.

Metode SMOTE adalah pendekatan over sampling sintetik menjadi banyak digunakan untuk permasalahan ini. Metode ini bekerja dengan cara membuat data sintesis pada kelas minoritas berdasarkan sampel yang ada, sehingga proporsi antar kelas menjadi seimbang. Metode ini dapat meningkatkan performa model. Dengan pendekatan ini, model cenderung lebih peka terhadap kelas minoritas dan mengurangi kesalahan terhadap pasien yang sebelumnya (Kumar dkk., 2022).

2.1.7 Cross-Validation

Cross-validation adalah teknik evaluasi model yang bertujuan untuk mengukur seberapa mampu model menggeneralisasi data baru yang belum pernah digunakan sebelumnya. Biasanya, teknik ini digunakan untuk menghindari kondisi model yang terlalu cocok dengan data pelatihan, namun gagal dalam memprediksi data baru secara akurat (*overfitting*). Teknik ini dapat memberikan estimasi pada suatu model yang bekerja pada data yang belum digunakan dalam proses pelatihan. Caranya adalah dengan membagi dataset menjadi bagian pelatihan dan validasi, selanjutnya model akan diuji pada berbagai subset data, sehingga dapat memberikan gambaran performa yang tidak bias (Marcot & Hanea, 2021).

Terdapat beberapa teknik cross validation yang umum digunakan, antar lain sebagai berikut :

1. *Holdout Validation*

Holdout Validation bekerja dengan cara membagi dataset menjadi dua bagian, yakni data pelatihan dan data pengujian. Teknik ini sederhana, namun kurang stabil, terutama pada dataset kecil.

2. *Leave-One-Out Cross-Validation (LOOCV)*

Leave-One-Out Cross-Validation bekerja dengan cara setiap satu per satu data digunakan untuk data pengujian, sisanya digunakan untuk data pelatihan. Teknik ini akurat, namun kurang efisien dalam segi komputasi.

3. *K-Fold Cross Validation*

K-Fold Cross Validation bekerja dengan cara membagi dataset menjadi k bagian (*folds*), kemudian proses pelatihan hingga pengujian dilakukan sebanyak k kali, dengan *fold* yang berbeda sebagai data uji pada tiap iterasi.

4. *Stratified K-Fold Cross Validation*

Stratified *K-Fold Cross Validation* adalah variasi dari *K-Fold Cross Validation*. Metode validasi silang ini bekerja dengan cara membagi dataset menjadi k bagian namun tetap mempertahankan proporsi distribusi kelas target pada setiap *fold*. Dengan metode ini, distribusi label dan komposisi label akan sebanding pada setiap *fold*. Sehingga metode validasi silang ini cocok diaplikasikan untuk data kalsifikasi yang tidak seimbang, karena dapat mengurangi bias dalam proses pelatihan dan evaluasi model.

Dari berbagai metode yang telah dijelaskan, metode *Stratified K-Fold Cross Validation* adalah metode yang sesuai untuk data yang tidak seimbang dalam konteks klasifikasi medis dengan data yang tidak seimbang untuk menghasilkan evaluasi model yang lebih stabil juga adil (T R dkk., 2023)

2.1.8 GridSearchCV

GridSearchCV merupakan metode pencarian sistematis terhadap kombinasi *hyperparameter* berdasarkan kriteria performa tertentu, seperti akurasi atau *F1-score*. Ruang parameter yang telah ditentukan sebelumnya, dieksplorasi secara menyeluruh, kemudian dilakukan juga cross validation pada setiap kombinasi parameter untuk mendapatkan hasil evaluasi yang stabil dan akurat. GridSearchCV cocok digunakan untuk eksplorasi awal dengan dataset kecil, karena jika diaplikasikan pada dataset yang besar dan kompleks, GridSearchCV membutuhkan waktu komputasi yang tinggi (Ali dkk., 2023).

2.1.9 Matriks Evaluasi

Matriks evaluasi adalah sebuah metode atau alat yang digunakan untuk mengukur seberapa baik model klasifikasi bekerja. Penting untuk memilih metrik evaluasi yang sesuai dengan kondisi dataset, misalnya distribusi kelas pada dataset, apakah distribusi datanya sudah seimbang atau belum, terutama pada jumlah ketidakseimbangan kelas antara jumlah kasus positif dan negatif, seperti pada data diagnosis penyakit. Hal ini dinilai

penting untuk menciptakan hasil prediksi yang tidak hanya akurat, tetapi juga mampu mendeteksi kasus positif secara tepat (Rainio dkk., 2024).

1. *Confusion Matrix*

Confusion matrix adalah sebuah tabel 2x2 yang digunakan untuk menunjukkan jumlah prediksi yang benar dan salah, baik untuk kelas positif maupun negatif, terdiri dari empat komponen *True Positive*, *True Negative*, *False Positive*, *False Negative*. Gambar 2.5 menampilkan *Confusion Matrix*.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Gambar 2. 5 Confusion Matrix

Dengan keterangan sebagai berikut :

TP (True Positive) : Jumlah data positif yang benar diprediksi positif oleh model.

TN (True Negative) : Jumlah data negatif yang benar diprediksi negatif oleh model.

FP (False Positive) : Jumlah data negatif yang salah diprediksi positif oleh model.

FN (False : Jumlah data positif yang salah diprediksi negatif Negative) oleh model.

Nilai akurasi, precision, recall dan f1-score yang menjadi indikator utama dalam untuk mengukur performa sebuah model terutama pada kasus klasifikasi diukur berdasarkan Confusion Matrix

2. Accuracy

Akurasi dihitung dari proporsi prediksi yang benar terhadap total data. Pada dataset yang tidak seimbang, model hanya memprediksi kelas mayoritas sehingga menghasilkan akurasi yang tinggi, sehingga akurasi cenderung tidak representatif untuk data yang tidak seimbang. Rumus akurasi ditampilkan pada Rumus (13).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

3. Precision

Precision dihitung dengan cara mengukur banyak prediksi positif yang benar-benar positif. Metrik ini sangat penting digunakan ketika konsekuensi dari false positif cukup besar. Precision dihitung dengan menggunakan Rumus (14).

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

4. Recall

Recall dihitung berdasarkan seluruh data positif yang berhasil dikenali model dari seluruh data positif yang ada. Dalam kasus diagnosis penyakit, metrik ini penting karena metrik ini menangkap semua kasus yang benar-benar positif. Recall dihitung dengan menggunakan Rumus (15).

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

5. F1-Score

F1-Score dihitung berdasarkan rata-rata harmonik dari precision dan recall. Metrik ini memberikan penilaian yang lebih seimbang pada kondisi data yang tidak seimbang. F1-Score dihitung dengan menggunakan Rumus (16).

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (16)$$

2.2 Penelitian Terdahulu

2.2.1 State of the Art

Pada bagian ini, akan dijelaskan rujukan yang digunakan dalam penelitian ini seperti yang dijelaskan pada Tabel 2.1.

Tabel 2. 1 *State of the Art*

No	Nama Peneliti/Journal		Hasil Penelitian
1	Peneliti (tahun)	(Chelliah dkk., 2024)	Penelitian ini berfokus pada prediksi PCOS dengan model machine learning yang dapat dijelaskan (Explainable
	Judul	Enhancing PCOS Prediction Using Machine	

No	Nama Peneliti/Journal		Hasil Penelitian
		Learning and Explainable AI	AI). Dataset yang digunakan adalah dataset PCOS yang bersumber dari UCI, dataset ini memuat variabel-variabel klinis seperti BMI, insulin, FSH, LH, dan rasio FSH/LH. Pada penelitian ini model XGBoost meraih akurasi tertinggi, dan hasil interpretasi dari SHAP menunjukkan bahwa BMI dan FSH/LH ratio merupakan fitur yang paling signifikan.
	Algoritma/ Metode	XGBoost, Decision Tree, Logistic Regression, SHAP (untuk interpretasi model)	
2	Peneliti	(Madhavi dkk., 2024)	Penelitian ini berfokus pada diagnosis PCOS dengan pembobotan fitur berbasis interpretabilitas. Dataset yang digunakan dalam penelitian ini adalah Dataset PCOS yang bersumber dari Kaggle dengan fitur-fitur yang mencakup hormon, berat badan, insulin, dan siklus menstruasi. Pada penelitian ini, Random Forest menghasilkan
	Judul	Detection of PCOS Leveraging Machine Learning Using Explainable AI	
	Algoritma/ Metode	Model klasifikasi : Random Forest, XGBoost, SVM Interpretasi : SHAP dan LIME	

No	Nama Peneliti/Journal		Hasil Penelitian
			akurasi tertinggi 94%, dan SHAP menunjukkan bahwa insulin dan LH/LSH ratio dominan.
3	Peneliti	(Gethsia dkk., 2024)	Penelitian ini berfokus pada identifikasi fitur kunci PCOS menggunakan <i>explainable AI</i> dalam model ensambel. Dataset yang digunakan adalah Dataset PCOS yang bersumber dari Kaggle, dengan fitur-fitur klinis seperti FSH/LH, berat badan, dan siklus menstruasi. Hasil penelitian menunjukkan bahwa SHAP mengidentifikasi LH dan insulin adalah faktor yang paling signifikan, ensambel model menghasilkan akurasi tertinggi hingga 93%.
	Judul	Interpreting PCOS Risk Factors Using Explainable AI and Ensemble Modeling Techniques.	
	Algoritma/ Metode	Random Forest, XGBoost, Stacking Classifier, SHAP	
4	Peneliti	(Akhtar dkk., 2023)	Penelitian ini berfokus pada peningkatan prediksi PCOS dengan Teknik oversampling dan ensambel learning. Dataset
	Judul	An Improved Prediction of PCOS Using	

No	Nama Peneliti/Journal		Hasil Penelitian
		SMOTE-based Oversampling and Stacking Classifier	yang digunakan adalah dataset PCOS yang bersumber dari UCI, dengan variable kunci
	Algoritma/Metode	SMOTE, Stacking (Random Forest, Logistic Regression, XGBoost)	variable hormonal dan siklus menstruasi. Hasil penelitian menunjukkan akurasi model meningkat dari 88% ke 94% setelah SMOTE dan <i>stacking classifier</i> diaplikasikan.
5	Peneliti	(Majumdar dkk., 2024)	Penelitian ini berfokus pada evaluasi berbagai
	Judul	Effectiveness of Class Imbalance Treatment for Diagnosis of PCOS Using Machine Learning Models : A Comparative Study	metode penanganan ketidakseimbangan kelas pada diagnosis PCOS. Dataset yang digunakan adalah dataset PCOS yang bersumber dari Kaggle yang mencakup data klinis dan hormonal. Hasil penelitian menunjukkan
	Algoritma/Metode	SMOTE, ADASYN. Random Undersampling, Random Forest, Decision Tree, SVM	SMOTE memberikan performa terbaik pada model Random Forest dengan akurasi 91%.

2.2.2 Matriks Penelitian

Ringkasan penelitian terdahulu yang memuat algoritma, fitur utama, dan metode yang digunakan sebagai pembanding pada penelitian ini tersaji pada Tabel 2.2.

Tabel 2. 2 Matriks Penelitian

Nama Penulis (Tahun)	Algoritma	Fitur			S M O T E	S H A P	Dataset Publik
		FSH/L H Ratio	Insulin	BMI			
(Chelliah dkk., 2024)	XGBoost, Decision Tree, Logistic Regression	√	√	√		√	√
(Madhavi dkk., 2024)	Random Forest, XGBoost, SVM	√	√			√	√
(Gethsia dkk., 2024)	Random Forest, XGBoost, Stacking Classifier	√	√	√		√	√
(Akhtar dkk., 2023)	Random Forest, Logistic Regression, XGBoost		√		√		√

Nama Penulis (Tahun)	Algoritma	Fitur			S M O T E	S H A P	Dataset Publik
		FSH/L H Ratio	Insulin	BMI			
(Majumdar dkk., 2024)	Random Forest, Decision Tree, SVM		√		√		√
(Vinothini dkk., 2024)	Random Forest, XGBoost, KNN, SVM(Vinothini dkk., 2024)	√	√	√	√		√
(Thakre, 2020)	Random Forest	√	√	√			√
(Khan Inan dkk., 2021)	XGBoost, SVM, Random Forest, Neural Networks	√		√	√		√
(Ahmed dkk., 2023; Panjwani dkk., 2025)	CNN, Random Forest, XGBoost	√	√	√		√	√
(Panjwani dkk., 2025)	Ensemble Learning, Random Forest			√		√	√

2.3.2 Penelitian Terdekat

Berdasarkan studi literatur yang telah dituangkan dalam *State of the Art* dan matriks penelitian, diperoleh informasi bahwa mayoritas penelitian menggunakan algoritma Random Forest dan XGBoost sebagai algoritma utama, karena kedua algoritma ini mampu menangani fitur kompleks dan non-linear. Selain itu, baik Random Forest maupun XGBoost banyak dikombinasikan dengan SHAP. Metode *balancing data* seperti SMOTE, ADASYN, dan *Random Undersampling* telah banyak digunakan, namun masih jarang yang mengombinasikannya dengan SHAP atau model interpretatif lainnya. Selain itu, penggunaan SHAP sudah digunakan untuk menjelaskan faktor dan risiko utama dari PCOS. Namun, penelitian yang mengombinasikan Random Forest atau XGBoost, SHAP, dan SMOTE masih jarang ditemukan. Hasil analisis dari SHAP sering menemukan bahwa fitur-fitur seperti FSH/LH ratio, insulin, dan BMI adalah fitur-fitur yang paling berkontribusi dalam diagnosis PCOS. Berdasarkan studi literatur yang telah dilakukan, terdapat celah penelitian yakni, belum ada penelitian yang menggabungkan teknik penyeimbangan data, algoritma ensambel, dan model interpretatif, padahal kombinasi ini penting untuk diaplikasikan guna memberikan informasi mengenai fitur-fitur penting yang berkontribusi pada diagnosis penyakit PCOS.