

BAB I

PENDAHULUAN

1.1 Latar Belakang

Polycystic Ovary Syndrome (PCOS) merupakan gangguan hormonal yang umum terjadi pada perempuan usia subur (Begum dkk., 2025). PCOS mulai diperkenalkan pada tahun 1935 oleh Irving F. Stein dan Michael L. Leventhal. Biasanya, PCOS menyerang perempuan dalam usia antara 15 hingga 44 tahun, terutama pada masa awal menstruasi hingga menjelang *menopause*. Ironisnya, perempuan baru menyadari dirinya terkena PCOS ketika akan memiliki keturunan, karena kondisi ini adalah salah satu faktor utama penghambat keturunan.

PCOS seringkali dikaitkan dengan penyakit reproduksi lainnya seperti Miom. Karena keduanya sama-sama dialami oleh perempuan pada usia subur dan dapat berakibat pada gangguan kesuburan. Padahal, kedua kondisi tersebut adalah dua kondisi yang berbeda. PCOS adalah penyakit gangguan hormonal yang ditandai dengan gejala siklus menstruasi yang tidak lancar, kenaikan berat badan, jerawat berlebihan, pertumbuhan yang tidak normal (*hirsutisme*), kelebihan hormon androgen, dan adanya kista pada ovarium (Johnson dkk., 2025). Sedangkan, miom (*uterine fibroids* atau *leiomyoma*) adalah tumor jinak yang terbentuk dari jaringan otot rahim. Meskipun pertumbuhan tumor dipengaruhi oleh hormon *estrogen*, miom bukan penyakit yang bersifat hormonal, melainkan struktural. Biasanya, miom ditandai dengan gejala nyeri

panggul, perdarahan atau menstruasi berat, hingga gangguan kesuburan (Sai & Mahaparale, 2024). Ukuran dan lokasi miom sangat menentukan gejala yang muncul.

PCOS meskipun dengan gejala yang cukup umum, sulit didiagnosis karena gejalanya mirip dengan gangguan hormonal lainnya, seperti *hipotiroidisme*, *hiperprolaktinemia*, dan gangguan adrenal (Bizuneh dkk., 2025; Joham dkk., 2025). Gejala PCOS seperti siklus menstruasi yang tidak teratur, jerawat berlebih, dan *hirsutisme* ditemukan juga pada penyakit *hipotiroidisme*, *hiperprolaktinemia*, dan gangguan adrenal. Pada *hipotiroidisme* ditandai dengan siklus menstruasi yang tidak lancar, kenaikan berat badan, dan kelelahan. Pertumbuhan rambut berlebih dan peningkatan androgen juga ditemukan pada gangguan adrenal yang disebabkan oleh hiperplasia adrenal kongenital (Yesiladali dkk., 2022). Siklus menstruasi yang tidak lancar dan infertilitas ditemukan juga pada *Hiperprolaktinemia* yang disebabkan oleh kadar prolaktin yang tinggi. Sehingga, seringkali menyebabkan kesalahan dalam diagnosis, terutama tanpa pemeriksaan hormonal dan ultrasonografi.

Kriteria Rotterdam 2003 adalah sebuah standar yang digunakan dalam diagnosis PCOS, hal ini membedakan PCOS dan penyakit hormonal lain (Zheng dkk., 2023). Dalam Kriteria Rotterdam, pasien didiagnosis PCOS apabila memenuhi setidaknya dua dari tiga kriteria: oligo atau anovulasi ditandai dengan siklus menstruasi yang tidak lancar, hiperandrogenisme yaitu ketidakseimbangan hormon androgen yang ditandai dengan pertumbuhan

rambut berlebih, jerawat yang parah di area dagu, jidat, yang terakhir adalah *polycystic ovary morphology* (PCOM) memerlukan pemeriksaan USG dan uji hormon ditandai dengan banyak folikel di sekitar ovarium, atau terdapat kadar *Anti Mullerian Hormone* (AMH) yang berlebih dan rasio *Luteinizing Hormone* (LH) dan *Follicle-Stimulating Hormone* (FSH) yang meningkat. Hal inilah yang menjadi gejala yang khas dalam diagnosis PCOS (van der Ham dkk., 2024).

Resistensi insulin merupakan salah satu penyebab utama terhadap timbulnya hiperandrogenisme dan gangguan ovulasi (Dutkiewicz dkk., 2024; Zeng dkk., 2020; Zheng dkk., 2023). Resistensi insulin menyebabkan hiperinsulinemia yang merangsang produksi androgen secara berlebihan oleh ovarium (Livadas dkk., 2022). Ditambah lagi obesitas dan gaya hidup yang tidak beraturan memperburuk kondisi ini, sehingga perempuan dengan kelebihan berat badan atau berisik diabetes tipe 2 memiliki kecenderungan tinggi untuk mengalami PCOS (Henney dkk., 2024).

Berdasarkan data dari WHO pada tahun 2023, umumnya sekitar 70% tidak terdeteksi, sehingga kemungkinan jumlahnya jauh lebih tinggi dibandingkan dengan yang tercatat (Syamsiah dkk., 2024). Di Indonesia, prevalensi PCOS juga mengalami peningkatan yang signifikan, dari 4-6% pada tahun 1990 menjadi 12-20% pada tahun 2023 (Sari dkk., 2023; Syamsiah dkk., 2024).

Deteksi dini pada PCOS menjadi sangat penting, karena jika tidak ditangani secara tepat, sekitar 53% Perempuan pengidap PCOS berisiko mengalami infertilitas, di samping 28% mengalami obesitas, 13% penyakit jantung, dan 6% menderita diabetes tipe 2 (Ajmal dkk., 2019; Syamsiah dkk., 2024). Kondisi ini tidak hanya menghambat kesuburan, melainkan dapat meningkatkan risiko diabetes tipe 2, obesitas, gangguan psikologis, dan penyakit jantung (Chen dkk., 2025; Muna Dinur Rafifah dkk., 2025). Di sisi lain, diagnosis PCOS secara klinis membutuhkan waktu dan biaya yang tidak sedikit. Pendekatan *machine learning* sebagai solusi untuk menghadapi tantangan dalam diagnosis dini PCOS sudah banyak digunakan untuk membangun model klasifikasi yang akurat dan efisien. Namun, hasil diagnosis seringkali bias karena ketidakseimbangan kelas pada dataset. Selain, itu model hanya mampu mendiagnosis tanpa mampu memberikan informasi mengenai fitur-fitur yang paling berkontribusi pada hasil diagnosis.

Berdasarkan studi literatur yang telah dilakukan, model Random Forest dan XGBoost merupakan salah satu model yang secara konsisten memberikan performa terbaik dalam kasus prediksi PCOS. Model Random Forest yang dikombinasikan dengan *Synthetic Minority Oversampling Technique* (SMOTE) mampu menghadapi tantangan yang umum pada dataset PCOS yaitu ketidakseimbangan data sehingga menghasilkan akurasi yang maksimal (Aprianti dkk., 2025; Permatasari Djaka & Winarsih, n.d.). Disisi lain, XGBoost juga berpotensi besar dalam klasifikasi data medis, termasuk PCOS. XGBoost dikenal karena memiliki performa yang unggul dalam mengatasi data

yang kompleks dan tidak seimbang, model XGBoost yang dikombinasikan dengan SMOTE untuk menghadapi ketidakseimbangan data PCOS menghasilkan akurasi yang cukup besar (Atika dkk., 2025).

Meskipun demikian, hasil prediksinya masih belum diketahui, karena hasil prediksi yang dihasilkan hanya fitur-fitur secara umum, tidak spesifik mengerucut pada fitur yang menjadi pemicu PCOS. Sehingga kondisi ini sukar dijelaskan secara langsung dan akan sulit diterima oleh tenaga medis. Untuk itu, diperlukan model yang tidak hanya mendiagnosis PCOS secara akurat, namun juga mampu menjelaskan mengapa hasil keputusan dari penggunaan model tersebut. Model *machine learning* yang dikombinasikan dengan *Explainable AI* seperti *Shapley Additive ExPlanations* (SHAP) mampu menghasilkan prediksi yang dapat dijelaskan fitur-fitur pemicu PCOS. Dalam konteks diagnosis PCOS, penggunaan SHAP dapat memberikan penjelasan yang akurat terkait kontribusi fitur pada setiap hasil prediksi, sehingga memperkuat kepercayaan terhadap penggunaan sistem. SHAP juga mampu mengidentifikasi fitur-fitur klinis yang penting seperti LH/FSH ratio, jumlah folikel, dan kadar AMH, yang dapat digunakan sebagai dasar pengambilan keputusan medis (G & P, 2024; Gethsia dkk., 2024).

Saat ini, belum ada penelitian yang secara komprehensif membandingkan model Random Forest dan XGBoost yang dikombinasikan dengan SMOTE dan SHAP untuk mendeteksi serta mengidentifikasi fitur-fitur yang berkontribusi terhadap risiko PCOS. Berdasarkan hal tersebut, penelitian ini memiliki tujuan untuk mengatasi tantangan prediksi PCOS dengan

membandingkan hasil interpretasi dari dua model *machine learning*, yaitu Random Forest dan XGBoost. Masing-masing model ini dikombinasikan dengan teknik *balancing data* yaitu SMOTE dan SHAP. Diharapkan kombinasi dari keempat metode ini mampu memberikan prediksi yang jauh lebih akurat, interpretatif, serta mampu mengatasi ketidak seimbangan data yang menjadi masalah umum pada dataset PCOS.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, diperoleh rumusan masalah sebagai berikut :

1. Bagaimana cara mengatasi ketidakseimbangan data pada dataset PCOS dalam penerapan model Random Forest dan XGBoost?
2. Bagaimana menghasilkan fitur-fitur yang berpengaruh dalam diagnosis PCOS menggunakan metode interpretasi model SHAP?
3. Bagaimana performa model Random Forest dan XGBoost setelah SMOTE dan SHAP diaplikasikan untuk diagnosis PCOS?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, penelitian ini memiliki tujuan sebagai berikut:

1. Menerapkan teknik SMOTE untuk mengatasi ketidakseimbangan data pada dataset PCOS dalam proses pelatihan model Random Forest dan XGBoost.

2. Mengimplementasikan metode SHAP untuk menginterpretasikan hasil diagnosis dari masing-masing model berdasarkan fitur-fitur yang paling berpengaruh terhadap prediksi PCOS dari nilai SHAP yang didapat.
3. Menganalisis performa model Random Forest dan XGBoost setelah dikombinasikan dengan SMOTE dan SHAP berdasarkan metrik evaluasi klasifikasi.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat, baik dari sisi akademis maupun praktis, diantaranya :

1. Menambah referensi ilmiah terkait penerapan *Explainable AI (XAI)* dalam model prediksi berbasis data medis.
2. Memberikan kontribusi pada pengembangan metode *machine learning* untuk prediksi penyakit, khususnya dalam konteks kesehatan reproduksi perempuan.
3. Memberikan gambaran awal mengenai fitur-fitur klinis yang paling berpengaruh terhadap risiko *Polycystic Ovary Syndrome* (PCOS).

1.5 Batasan Masalah

Beberapa batasan yang diterapkan dalam penelitian ini, agar fokus dan ruang lingkup penelitian ini jelas dan terarah, adalah sebagai berikut:

1. Dataset yang digunakan bersumber dari platform Kaggle sebagai dataset publik yang tersedia secara daring, bukan data klinis langsung dari instansi medis atau rumah sakit.

2. Penelitian ini hanya menggunakan algoritma Random Forest dan XGBoost untuk mendiagnosis dini penyakit PCOS.
3. Teknik balancing data yang digunakan terbatas pada SMOTE.
4. Pendekatan *Explainable AI* yang diterapkan terbatas pada SHAP, tanpa membandingkan dengan pendekatan interpretasi lain.
5. Evaluasi performa model dilakukan menggunakan metrik klasifikasi yaitu akurasi, *precision*, *recall*, dan *F1-score*, tanpa validasi klinis lebih lanjut.