

BAB I

PENDAHULUAN

Bab ini menyajikan mengenai latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam dokumen skripsi yang berjudul *Prediksi Polycystic Ovary Syndrome (PCOS) Menggunakan Metode Ensemble Learning*.

1.1. Latar Belakang

Polycystic Ovary Syndrome (PCOS) merupakan salah satu gangguan hormon yang paling umum terjadi pada wanita usia reproduktif. PCOS ditandai dengan ketidakseimbangan hormon seksual pada tubuh wanita yang dapat mempengaruhi kesuburan serta kesehatan metabolisme (Oğuz dkk., 2024). Gejala yang terkait dengan PCOS antara lain adalah siklus menstruasi yang tidak teratur, terlambat, atau bahkan tidak hadir, serta peningkatan risiko berbagai komplikasi kesehatan seperti diabetes tipe II, diabetes gestasional, pertumbuhan rambut berlebih (hirsutisme), jerawat, obesitas, dan tingginya kadar testosteron. PCOS juga diketahui dapat meningkatkan risiko hipertensi, depresi, dan masalah kesehatan lainnya.

Sistem reproduksi wanita terdiri dari ovarium, tuba falopi, rahim, dan vagina. Ovarium menyimpan sel telur dalam kantung kecil berisi cairan yang disebut folikel. Aktivitas ovarium diatur oleh hormon yang dihasilkan kelenjar pituitari, yang terletak di dasar otak. Setiap bulan, kelenjar ini melepaskan hormon perangsang folikel (FSH) dan hormon *luteinizing* (LH) ke dalam aliran darah untuk memicu ovarium mematangkan sel telur. Saat kadar LH meningkat, folikel melepaskan estrogen, yang mengakibatkan ovulasi. Namun, pada kasus PCOS, proses ini dapat terganggu, menyebabkan gangguan pada siklus menstruasi dan meningkatkan risiko infertilitas (Hajam dkk., 2023).

Menurut *World Health Organization (WHO)* pada tahun 2023, PCOS diperkirakan mempengaruhi sekitar 8% sampai dengan 13% wanita usia reproduktif

di seluruh dunia, dan hingga 70% kasus PCOS tidak terdiagnosis. Meningkatnya angka kejadian PCOS dalam beberapa tahun terakhir menjadikan kebutuhan untuk deteksi dini dan intervensi lebih awal semakin mendesak. PCOS biasanya dideteksi melalui kombinasi metode seperti pemeriksaan ultrasonografi (USG), analisis kadar hormon dalam darah, dan evaluasi gejala klinis. Namun, prediksi PCOS menjadi sulit karena sering kali membutuhkan pemeriksaan laboratorium yang kompleks dan memakan waktu, sehingga memperlambat proses diagnosis. Oleh karena itu, diperlukan sistem deteksi dini yang dirancang untuk membantu para ahli dalam memprediksi PCOS secara lebih mudah dan cepat, sehingga memungkinkan pengambilan keputusan yang lebih efisien tanpa harus sepenuhnya bergantung pada pemeriksaan konvensional. Dengan penerapan pembelajaran mesin, hasil deteksi dapat langsung diperoleh secara otomatis tanpa memerlukan penerjemahan khusus dari ahli, sehingga lebih efisien sebagai alternatif deteksi dini PCOS.

Penelitian oleh Denny dkk. (2019) ,menunjukkan bahwa berbagai metode klasifikasi dapat digunakan untuk memprediksi PCOS. Dalam penelitian tersebut, algoritma *Random Forest* adalah model paling akurat berdasarkan akurasi yakni 89,02%. Sedangkan dalam penelitian lain oleh Rahman dkk. (2024) mengeksplorasi penggunaan pembelajaran mesin dalam memprediksi PCOS, dan membuktikan bahwa *Random Forest* terbukti memberikan performa terbaik, mencapai *precision* sebesar 96,67%, *recall* 82,86% dan akurasi 94%.

Meskipun metode-metode ini menunjukkan hasil yang baik, penggunaan *Ensemble Learning* dapat lebih meningkatkan akurasi prediksi. Salah satu penelitian yang dilakukan oleh Cendani & Wibowo (2022) menunjukkan bahwa metode *Ensemble Learning* dapat diterapkan untuk memprediksi penyakit diabetes serta meningkatkan akurasi prediksi. Dalam penelitian ini, tiga metode *ensemble*, yaitu *Bagging*, *Boosting*, dan *Stacking*, diuji dan dibandingkan. Hasilnya, setiap metode mencapai akurasi yang lebih tinggi dibandingkan model individual. Dimana akurasi model individual *Random Forest* menghasilkan akurasi 79,87% . Sedangkan setelah dilakukan *Ensemble Learning*, akurasinya meningkat menjadi 81,82%. Hasil ini menunjukkan bahwa *Ensemble Learning* dapat meningkatkan performa prediksi dan memberikan hasil yang lebih andal.

Penelitian lain juga dilakukan oleh Alhamad dkk. (2019) dimana metode *Ensemble Learning* juga dapat digunakan dalam memprediksi penyakit jantung. Hasil penelitian menunjukkan bahwa model *Ensemble Learning* mencapai akurasi 85,21%, melebihi beberapa metode individu seperti *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), C4.5, *Naïve Bayes*, dan K-Nearest Neighbor dengan akurasi berturut-turut 78,07%, 82,44%, 84,87%, dan 81,26%. Berdasarkan penelitian tersebut, pendekatan *ensemble* terbukti mampu mengatasi keterbatasan model tunggal, menghasilkan prediksi yang lebih andal dan efisien dalam mendeteksi PCOS.

Pemanfaatan pembelajaran mesin telah terbukti efektif dalam memprediksi PCOS melalui berbagai model, seperti *Random Forest*, *Decision Tree*, *Support Vector Machine*, dan *Adaboost*. Namun, model-model ini masih memiliki kelemahan, terutama dalam hal sensitivitas dan spesifisitas, yang dapat menyebabkan kesalahan dalam mengidentifikasi individu dengan PCOS sebagai tidak terdiagnosis dan mengklasifikasikan individu tanpa PCOS sebagai terdiagnosis.

Oleh karena itu, dilakukan penelitian ini untuk menganalisis data pasien yang terdiagnosis dengan dan tanpa PCOS serta menggunakan matrik evaluasi seperti akurasi, *sensitivity*, *specificity*, *precision*, *recall*, dan *F1-score* untuk membandingkan kinerja model *Random Forest*, *Artificial Neural Network* (ANN), dan *Ensemble Learning*. Penelitian diharapkan dapat menunjukkan bahwa metode *Ensemble Learning* memiliki akurasi yang lebih tinggi dibandingkan model *individual*, sehingga menjadi pilihan efektif untuk deteksi dini PCOS. Dengan kontribusi ini, diharapkan penelitian ini dapat mendukung pengembangan teknologi diagnostik yang lebih akurat dan andal, sehingga membantu pasien PCOS untuk mendapatkan perawatan lebih awal dan mencegah komplikasi yang lebih serius.

1.2. Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana penerapan metode *Ensemble Learning* dapat meningkatkan akurasi, *sensitivity*, *specificity*, *precision*, *recall*, dan *F1-score* model dalam memprediksi status PCOS. Penelitian ini berfokus pada evaluasi efektivitas pendekatan *Ensemble Learning*, yang menggabungkan beberapa model prediksi, dalam meningkatkan akurasi deteksi status PCOS dibandingkan dengan pendekatan menggunakan model individual.

1.3. Tujuan dan Manfaat Penelitian

Penelitian ini bertujuan untuk memahami sejauh mana pendekatan *Ensemble Learning*, yang menggabungkan beberapa model prediksi, dapat meningkatkan akurasi, *sensitivity*, *specificity*, *precision*, *recall*, dan *F1-score* dari prediksi dibandingkan dengan model individual. Selain itu, penelitian ini juga menganalisis pengaruh pemilihan data dan kontribusi atribut terhadap hasil prediksi yang diperoleh, sehingga dapat memberikan manfaat dalam pengembangan teknologi diagnostik yang lebih akurat untuk deteksi dini status PCOS.

Manfaat dari penelitian ini diharapkan dapat memberikan kontribusi dalam mempermudah masyarakat awam melakukan deteksi dini PCOS, sehingga dapat mendukung upaya pencegahan dan pengelolaan kondisi ini secara lebih efektif. Selain itu, hasil penelitian ini dapat menjadi referensi bagi pengembangan sistem prediksi berbasis pembelajaran mesin untuk gangguan kesehatan lainnya. Deteksi dini dengan metode ini membantu pasien mendapatkan perawatan lebih awal, sehingga dapat mencegah komplikasi serius.

1.4. Ruang Lingkup

Penelitian ini memerlukan adanya ruang lingkup agar penelitian yang dilakukan lebih terarah dan mencapai sasaran yang diharapkan. Ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini diambil dari *platform* Kaggle dan mencakup semua parameter fisik dan klinis yang relevan untuk menentukan masalah PCOS dan infertilitas.
2. Penelitian ini akan difokuskan pada analisis dan pengembangan model pembelajaran mesin dengan menggunakan metode *Stacking* yang menggabungkan model *Random Forest* dan *Artificial Neural Network* (ANN).
3. Cakupan penelitian ini terbatas pada data individu yang mewakili populasi umum tanpa pembatasan kelompok geografis tertentu, sehingga hasil penelitian diharapkan bersifat general dan dapat diaplikasikan pada populasi yang lebih luas.

1.5. Sistematika Penulisan

Sebagai gambaran mengenai pembahasan penyusunan laporan skripsi ini secara urut dan jelas, maka dibentuk sistematika sebagai berikut:

BAB I Pendahuluan

Bab ini menyajikan latar belakang, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup, serta sistematika penulisan yang digunakan dalam dokumen skripsi yang berjudul Prediksi *Polycystic Ovary Syndrome* (PCOS) Menggunakan Metode *Ensemble Learning*.

BAB II LANDASAN TEORI

Bab ini memuat tinjauan pustaka dan teori-teori dasar yang digunakan dalam skripsi ini. Bab ini terdiri atas beberapa sub-bab yang mencakup *state of the art*, konsep dasar pembelajaran mesin, algoritma prediksi yang relevan untuk kasus PCOS, metode *stacking*, serta kajian literatur yang mencakup studi-studi sebelumnya terkait prediksi kondisi kesehatan menggunakan model *Random Forest*, *Artificial Neural Network* (ANN) dan metode *Ensemble Learning*.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan rancangan penelitian dan tahapan-tahapan yang dilakukan dalam skripsi Prediksi *Polycystic Ovary Syndrome* (PCOS) Menggunakan Metode *Ensemble Learning*. Tahapan penelitian meliputi pengumpulan data, *Exploratory Data Analysis* (EDA), pra-pemrosesan data, seleksi fitur, pembagian dataset, *Random Forest*, *Artificial Neural Network* (ANN), *Ensemble Stacking*, optimasi model, serta evaluasi performa model berdasarkan *Confusion Matrix*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dan analisis yang dilakukan sesuai dengan metodologi penelitian. Pembahasan mencakup performa model prediksi PCOS yang dikembangkan, perbandingan akurasi model individual (*Random Forest* dan ANN) dengan metode *stacking*, interpretasi hasil model, efektivitas

model prediksi, serta pembahasan terkait kontribusi masing-masing atribut dalam memprediksi status PCOS.

BAB V PENUTUP

Bab ini berisi kesimpulan dari bab-bab yang telah dijelaskan sebelumnya dan saran untuk penelitian lebih lanjut.