

ABSTRAK

Polycystic Ovary Syndrome (PCOS) merupakan gangguan hormonal kompleks yang memengaruhi perempuan usia subur dan seringkali sulit didiagnosis secara dini. Tantangan utama dalam diagnosis PCOS menggunakan machine learning terletak pada ketidakseimbangan data dan keterbatasan interpretabilitas model. Penelitian ini bertujuan untuk menganalisis fitur-fitur penting dalam diagnosis PCOS dengan membandingkan dua algoritma machine learning, yaitu Random Forest dan XGBoost. Kedua model dikombinasikan dengan teknik penanganan data imbalance Synthetic Minority Over-sampling Technique (SMOTE) serta metode interpretasi model Shapley Additive Explanations (SHAP). Dataset yang digunakan berasal dari platform Kaggle, terdiri dari 541 data pasien dan 45 fitur klinis. Proses penelitian meliputi preprocessing data, eksplorasi data, pelatihan model dengan GridSearchCV dan Stratified K-Fold Cross-Validation, evaluasi performa model menggunakan metrik akurasi, precision, recall, dan F1-score, serta interpretasi hasil dengan SHAP. Hasil penelitian menunjukkan bahwa model XGBoost dengan kombinasi SMOTE dan SHAP memberikan performa terbaik dalam klasifikasi dan interpretasi fitur diagnosis PCOS. Fitur Follicle No. (R), Hair growth (Y/N), dan Skin darkening (Y/N) muncul sebagai fitur yang paling berpengaruh secara konsisten. Pendekatan ini memberikan gambaran yang akurat dan dapat dipahami secara medis, sehingga mendukung proses diagnosis yang lebih transparan dan efektif.

Kata Kunci — *Explainable AI, PCOS, Random Forest, SHAP, XGBoost*