

## INTISARI

### KLASIFIKASI PCOS MENGGUNAKAN MODEL STACKING ENSEMBLE LEARNING DENGAN SMOTE-ENN

Oleh :  
Silmi Utami Putri  
24/549092/PPA/06935

*Polycystic Ovary Syndrome (PCOS)* merupakan gangguan endokrin kompleks yang dapat menyebabkan komplikasi serius seperti infertilitas dan gangguan metabolik. Penelitian ini menggunakan dataset gabungan (*merge dataset*) dari repositori publik Kaggle yang mencakup total 2.541 data pasien, hasil integrasi dari tiga sumber utama: data pasien tanpa infertilitas, data infertilitas, dan dataset extended dari 10 rumah sakit di India. Dataset ini memiliki ketimpangan kelas signifikan dengan rasio 2,24:1, di mana kelas minoritas (PCOS) hanya 30,89% dibandingkan kelas mayoritas (Non-PCOS) sebesar 69,11%, yang berisiko menyebabkan model bias dan gagal mendeteksi kasus positif (*false negative*).

Untuk mengatasi kendala tersebut, penelitian ini menerapkan metode *Stacking Ensemble* yang menggabungkan *Random Forest*, *AdaBoost*, dan *CatBoost* dengan *Logistic Regression* sebagai *meta-learner*, dikombinasikan dengan teknik hybrid SMOTE-ENN untuk menangani ketidakseimbangan kelas. Proses SMOTE-ENN berhasil menyeimbangkan distribusi data latih menjadi 49,42% vs 50,58% serta membersihkan *noise data*. Model *baseline* menggunakan *Logistic Regression* dan *Naive Bayes* dilatih pada data asli sebagai pembandingan.

Hasil evaluasi menunjukkan metode yang diusulkan mencapai peningkatan performa dibandingkan model *baseline* pada data *testing* dan penelitian sebelumnya, dengan akurasi 99%, *precision* 99%, *recall* 99%, dan F1-Score 99%. Pencapaian ini membuktikan bahwa kombinasi *Stacking Ensemble* dengan SMOTE-ENN mampu melakukan klasifikasi medis yang akurat dan dapat diandalkan untuk diagnosis dini PCOS dalam praktik klinis.

**Kata Kunci:** PCOS, *Machine learning*, *Stacking*, SMOTE-ENN, *Imbalanced data*.

## ABSTRACT

### PCOS CLASSIFICATION USING STACKING ENSEMBLE LEARNING WITH SMOTE-ENN

by:

Silmi Utami Putri  
24/549092/PPA/06935

*Polycystic Ovary Syndrome* (PCOS) is a complex endocrine disorder that can lead to serious complications, such as infertility and metabolic disturbances. This study utilizes a merged dataset from public Kaggle repositories, comprising a total of 2,541 patient records integrated from three primary sources: non-infertility patient data, infertility data, and an extended dataset from 10 hospitals in India. The dataset exhibits a significant class imbalance with a ratio of 2.24:1, where the minority class (PCOS) accounts for only 30.89% compared to the majority class (Non-PCOS) at 69.11%. This imbalance poses a risk of model bias and potential failure in detecting positive cases (false negatives).

To address these challenges, this research implements a Stacking Ensemble method combining Random Forest, AdaBoost, and CatBoost with Logistic Regression as the meta-learner integrated with the SMOTE-ENN hybrid technique to handle class imbalance. The SMOTE-ENN process successfully balanced the training data distribution to 49.42% vs. 50.58% while effectively removing data noise. Baseline models using Logistic Regression and Naive Bayes were trained on the original data for comparative analysis.

Evaluation results demonstrate that the proposed method achieved a significant performance increase compared to baseline models and previous studies, reaching 99% accuracy, 99% precision, 99% recall, and a 99% F1-Score. These findings prove that the combination of Stacking Ensemble and SMOTE-ENN provides accurate and reliable medical classification for the early diagnosis of PCOS in clinical practice.

**Keywords:** PCOS, Machine learning, Stacking, SMOTE-ENN, Imbalance data.