PCOS Prediction Project Report

Objective

The goal of this project is to predict the likelihood of Polycystic Ovary Syndrome (PCOS) based on various clinical, physical, and lifestyle-related features using machine learning and exploratory data analysis (EDA). The project emphasizes data-driven medical decision support using interpretable and accurate models.

Dataset Overview

The dataset contains multiple features such as:

- Demographic features: Age, Height, Weight, Blood Group
- Clinical & Lifestyle features: Menstrual Cycle, Skin Darkening, Hair Loss, Fast Food Consumption, etc.
- Target variable: PCOS Diagnosis (1 = PCOS, 0 = Non-PCOS)

Exploratory Data Analysis (EDA)

Step 1: Correlation Heatmap

Why: To identify relationships between features and the target variable.

Findings:

- PCOS Diagnosis is positively correlated with Menstrual Cycle Interval (0.44), Excessive Facial/Body Hair, Skin Darkening, Weight Gain.
- Negatively correlated with Regular Periods (-0.54).
- Some features like Blood Group, Hair Loss, and Exercise show no significant correlation.

Feature Importance Analysis

Step 2: Permutation Feature Importance

Why: To quantify the effect of each feature on the model's prediction.

Top Influential Features:

1. Regular Periods
2. Menstrual Cycle Interval
3. Skin Darkening

PCOS Prediction Project Report

Step 3: Class-wise Feature Importance

For PCOS (Class 1): Regular Periods, Menstrual Cycle, Skin Darkening, Weight are strong indicators.

For Non-PCOS (Class 0): Feature importance is relatively flat.

Partial Dependence Plot (PDP)

Step 4: PDP on Menstrual Cycle Interval

Why: To visualize the marginal effect of a feature on prediction probability.

Insight: Longer or irregular menstrual cycles directly increase the likelihood of PCOS prediction.

Machine Learning Models

Step 5: Model Training & Comparison

Why: To identify the best algorithm for PCOS prediction.

Top Performing Model: Stacking (SVM + Logistic Regression) with F1 Score = 0.71.

Simpler models like Decision Tree underperform.

Final Insights & Conclusion

Key indicators of PCOS:

- Irregular or absent periods
- Abnormal menstrual cycle interval
- Weight gain, hirsutism, skin pigmentation

Less relevant features: Blood group, exercise, acne, hair loss

Stacking models outperform single models. SVM + LR stacking is recommended for deployment.

Future Scope

PCOS Prediction Project Report

Add hormonal test data to improve accuracy. Apply SHAP or LIME for deeper interpretability. Explore Power BI dashboard for clinician-friendly visual analytics.