PCOS (Polycystic Ovary Syndrome)

Meghnad Saha Institute of Technology
Kolkata, India
www.msit.edu.in

***Abstract*—Polycystic Ovary Syndrome (PCOS) is a common endocrine disorder that affects women of reproductive age. It is a complex condition characterized by hormonal imbalances, metabolic disturbances, and reproductive issues. PCOS is one of the leading causes of female infertility and can have significant impacts on a women's overall health and well-being.**

Keywords—PCOS, Women, Polycystic

I. Reasons

Polycystic ovary syndrome (PCOS) is a hormonal disorder that affects some people who have ovaries. PCOS can cause different symptoms, such as irregular or missed periods, heavy bleeding during periods, acne or excessive hair growth due to high levels of androgens, and enlarged or cystic ovaries. However, not everyone with PCOS has the same symptoms or develops cysts on their ovaries.

II. Analysis

A. Observation from Box Plot:

The boxplot reveals some insights about the hormonal disorder PCOS, which affects some people who have ovaries. One of the patterns is that patients who have more follicles are more likely to have PCOS than those who have fewer follicles. Another pattern is that patients who have shorter menstrual cycles are more prone to PCOS than those who have longer cycles. The boxplot also shows that the average age of patients who are diagnosed with PCOS is 30 years old.

B. Observation from Bar Plot:

The bar plot illustrates some factors that are associated with PCOS, a hormonal disorder that affects some people who have ovaries. Some of the factors are weight gain, hair growth, pimples, hair loss, fast food consumption and skin darkening. These factors indicate a higher chance of having PCOS. The bar plot also shows that 33 % of the patients in our dataset have PCOS syndrome. Moreover, the bar plot reflects the current research that PCOS increases the risk of endometrial cancer in women of all ages, but not ovarian or breast cancer. This suggests that PCOS may lead to gynecological cancer complications.

III. Model Prediction

PCOS is a hormonal disorder that can be diagnosed using prediction models. Prediction models are methods that use machine learning and data mining to analyze non-invasive features such as age, weight, cycle length, follicle number and hormonal levels. Prediction models should follow standardized criteria and metrics to ensure their reliability and validity. Prediction models can also benefit patients who have PCOS by helping them to identify their risk, receive early treatment and understand their condition better.

A. Logistic Regression Model:

PCOS can be predicted using a dataset, which implements a machine learning model applying logistic regression. The target variable "PCOS (Y/N)" is kept distinct from the input characteristics "X" and saved in a variable called "pcos" along with the dataset. Using the 'train_test_split' function, the data is split into training and testing sets after the input features are scaled using Scikit-learn's MinMaxScaler. Using the 'fit' technique, a logistic regression model is generated.

a) Confusion Matrix: We create a confusion matrix for evaluating the performance of the model. The true positives, true negatives, false positives, and false negatives are all presented thoroughly within this matrix. Using Seaborn, a heatmap is produced to improve the visual interpretation of the confusion matrix. This improves in identifying trends and provides a better understanding of the model's accuracy.

B. Decision Tree Model:

The Decision Tree Classifier class is executed with the random_state parameter set to 99 for reproducibility and the requirement parameter set to 'entropy' for dividing the data. The trained model is loaded into the loaded_model_dec_tree variable. The accuracy_score function from Scikit-learn is used to determine the model's accuracy score. The model is used for predicting target values for the test data (X_test). The classification report creates a thorough report that is printed to the console and includes the precision, recall, F1-score, and support for each class.

a) Confusion Matrix: Like the logistic regression model, the confusion matrix is calculated using the Scikit-learn confusion matrix tool. The confusion matrix is subsequently represented visually by a heatmap produced with Seaborn.

C. Random Forest Model:

A random forest classifier with a 9-n_estimators' parameter by importing the Random Forest Classifier class from Scikit-learn's collection of modules. The fit method is used to train the model, which is then saved in the rf variable. The accuracy score is then calculated by using the Scikit-learn accuracy_score function to forecast target values for the test data.

a) Confusion Matrix: Similar to the other models, the confusion matrix is calculated using the Scikit-learn confusion_matrix function. The confusion matrix is subsequently represented visually by a heatmap produced with Seaborn.

IV. Conclusion

PCOS is one of the most common disorders affecting women of reproductive age. As a syndrome, it has multiple components, including reproductive, metabolic, and cardiovascular, with long-term health concerns that cross the life span. It's important to note that medical knowledge is constantly evolving, and there may have been new developments in PCOS research and treatment beyond my last update in September 2021. Therefore, it's best to consult with a healthcare professional or a gynecologist for the most current and personalized information regarding PCOS.

References

[1] Karnika Kapoor, "PCOS Diagnosis,"