POLYCYSTIC OVARY SYNDROME CLASSIFICATION USING MACHINE LEARNING

ANINDUSTRYORIENTEDMINIREPORT

Submitted to

JAWAHARLALNEHRUTECNOLOGICALUNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOROFTECHNOLOGY

In

COMPUTERSCIENCEAND ENGINEERING

Submitted By

PUTTA GRESILA **21UK1A05G2**

Under the guidance of

Dr. J. Srikanth

Assistant Professor

DEPARTMENTOFCOMPUTERSCIENCEANDENGINEERING VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S)

506005

DEPARTMENTOF

COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE (WARANGAL)

CERTIFICATE OF COMPLETION
INDUSTRY ORIENTED MINIPROJECT

This is to certify that the Mini-Project entitled "POLYCYSTIC OVARY SYNDROME CLASSIFICATION USING MACHINE LEARNING" is being submitted by PUTTA GRESILA (21UK1A05G2), in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2024.

Project Guide
Dr. J. Srikanth
(Professor)

HOD
Dr. Naveen Kumar Rangaraju
(Professor)

External

ACKNOWLEDGEMENT

I wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr SYED MUSTHAK AHMED,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this Mini Project in the institute.

I extend my heartfelt thanks to **Dr. NAVEEN KUMAR RANGARAJU**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the Mini-PROJECT.

I express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the Mini-Project and for their support in completing the Mini-Project.

I express heartfelt thanks to the guide, **Dr. J. Srikanth,** professor, Department of CSE for his constant support and giving necessary guidance for completion of this Mini-Project.

Finally, I express my sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

PUTTA GRESILA **(21UK1A05G2)**

ABSTRACT

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder among women of reproductive age, characterized by a range of symptoms including irregular menstrual cycles, hyperandrogenism, and polycystic ovaries. The complexity and variability of PCOS symptoms make its diagnosis challenging, necessitating advanced diagnostic tools. This study explores the application of machine learning techniques for the classification and diagnosis of PCOS. By leveraging various machine learning algorithms, including Decision Trees, Support Vector Machines, and Neural Networks, we aim to develop a robust predictive model based on clinical and biochemical features. The dataset comprises medical records and diagnostic indicators from diverse patient profiles, ensuring comprehensive analysis and model generalizability. Our results demonstrate significant improvements in diagnostic accuracy and efficiency, highlighting the potential of machine learning in enhancing PCOS detection and personalized treatment strategies. This research underscores the importance of integrating artificial intelligence in healthcare for better management and understanding of complex medical conditions like PCOS.

TABLE OF CONTENTS:-

1. INTRODUCTION

OVERVIEW

Polycystic Ovary Syndrome (PCOS) is a multifaceted endocrine disorder affecting a significant percentage of women of reproductive age. Characterized by symptoms such as irregular menstrual cycles, excessive androgen levels, and the presence of ovarian cysts, PCOS can lead to long-term health issues like diabetes, cardiovascular disease, and infertility. Given the heterogeneity of the symptoms, accurate and timely diagnosis is often complex and challenging. Machine learning (ML) offers promising solutions to these challenges by leveraging data-driven approaches to enhance diagnostic accuracy and predict potential complications. Machine learning involves training algorithms to recognize patterns and make decisions based on data. In the context of PCOS classification, ML algorithms are trained using clinical, biochemical, and ultrasonographic data from patients. The primary goal is to develop models that can accurately classify whether a patient has PCOS and, in some cases, predict the severity of the condition. Machine learning holds significant promise in revolutionizing the diagnosis and management of Polycystic Ovary Syndrome. By leveraging advanced algorithms and diverse datasets, ML can improve diagnostic accuracy, personalize treatment plans, and ultimately enhance patient outcomes. As research progresses, the integration of AI in clinical practice will likely become more prevalent, offering new avenues for understanding and managing PCOS.

PURPOSE

The purpose of classifying Polycystic Ovary Syndrome (PCOS) using machine learning is to enhance the accuracy, efficiency, and consistency of diagnosing this complex and heterogeneous endocrine disorder. Traditional diagnostic methods for PCOS are often

subjective, time-consuming, and can vary significantly among healthcare providers due to the multifaceted nature of the condition and its diverse manifestations. Machine learning offers a data-driven approach to overcome these challenges by analyzing large datasets of clinical, biochemical, and imaging features to identify patterns and correlations that may not be apparent through conventional methods.

Specifically, the objectives of this purpose are:

1. **Improve Diagnostic Accuracy**: Develop robust machine learning models that can accurately classify individuals as having PCOS or not, based on a comprehensive set of clinical and biochemical markers. This can help reduce misdiagnosis and ensure that patients receive appropriate treatment in a timely manner.
2. **Enhance Early Detection**: Utilize predictive analytics to identify early signs and risk factors associated with PCOS, facilitating earlier intervention and management. Early detection can help mitigate long-term health complications related to PCOS, such as diabetes and cardiovascular diseases.
3. **Personalize Treatment Plans**: Leverage machine learning to stratify patients based on the severity and specific characteristics of their PCOS. This stratification can aid in tailoring personalized treatment plans that address the unique needs of each patient, improving clinical outcomes and patient satisfaction.
4. **Standardize Diagnostic Criteria**: Create a standardized, objective framework for PCOS diagnosis that can be widely adopted across different clinical settings. This can help reduce variability in diagnosis and ensure that all patients receive consistent and evidence-based care.
5. **Advance Research and Understanding**: Facilitate deeper insights into the underlying mechanisms and progression of PCOS by analyzing large and diverse datasets. This can contribute to the development of new therapeutic targets and interventions, further advancing the field of PCOS research.

2. LITERATURE SURVEY

EXISTINGPROBLEM

Polycystic Ovary Syndrome (PCOS) is a complex and multifaceted endocrine disorder that affects approximately 10% of women of reproductive age. Despite its prevalence, the diagnosis and management of PCOS remain challenging due to several critical issues:

1. Heterogeneous Presentation: PCOS manifests through a broad spectrum of symptoms, including irregular menstrual cycles, hyperandrogenism (excessive male hormones), and polycystic ovaries, which vary significantly among individuals. This heterogeneity complicates the development of a one-size-fits-all diagnostic criterion.

2. Lack of Standardized Diagnostic Criteria: There are multiple diagnostic criteria for PCOS (e.g., Rotterdam, NIH, AE-PCOS), leading to inconsistencies in diagnosis. These criteria often rely on subjective clinical judgment and can result in variability in diagnosis across different healthcare providers and settings.

3. Underdiagnosis and Misdiagnosis: Due to the overlapping symptoms of PCOS with other conditions (e.g., hypothyroidism, hyperprolactinemia), and the reliance on subjective assessments, PCOS is frequently underdiagnosed or misdiagnosed. This leads to delays in appropriate treatment and management.

4. Limited Understanding of Etiology: The exact cause of PCOS remains unclear, with genetic, environmental, and lifestyle factors all playing potential roles. This limited understanding hinders the development of targeted therapies and effective prevention strategies.

5. Inadequate Patient Management: Current management strategies for PCOS often involve a trial-and-error approach due to the variability in patient response to treatment.

This can result in suboptimal care, increased patient frustration, and poor long-term health outcomes.

6. Long-term Health Implications: Women with PCOS are at an increased risk for several long-term health issues, including type 2 diabetes, cardiovascular disease, infertility, and mental health disorders. Early and accurate diagnosis is crucial for preventing these complications, but existing diagnostic approaches are not sufficiently reliable.

PROPOSED SOLUTION

To address the challenges associated with the diagnosis and management of Polycystic Ovary Syndrome (PCOS), we propose the development and implementation of a machine learning-based classification system. This solution aims to enhance diagnostic accuracy, standardize criteria, and facilitate personalized treatment plans. The key components of the proposed solution include:

- **Data Collection and Integration**

  - **Comprehensive Dataset**: Gather extensive data from diverse sources, including clinical records, biochemical tests, ultrasound imaging, and patient-reported outcomes. Ensure that the dataset is representative of various demographics and PCOS phenotypes.
  - **Feature Extraction**: Identify and extract relevant features from the collected data, such as hormonal levels, menstrual cycle patterns, ultrasound characteristics, and clinical symptoms.

- **Preprocessing and Data Cleaning**

  - **Data Normalization**: Normalize the data to ensure consistency across different units and scales.
  - **Handling Missing Data**: Implement techniques for dealing with missing or incomplete data, such as imputation or exclusion methods.
  - **Feature Selection**: Use statistical and machine learning methods to select the most informative features that contribute to the accurate classification of PCOS.

- **Machine Learning Model Development**

  - **Algorithm Selection**: Evaluate and select appropriate machine learning algorithms for classification, such as Decision Trees, Support Vector Machines (SVM), Random Forests, and Neural Networks. Ensemble methods can also be explored to combine the strengths of multiple algorithms.
  - **Model Training**: Train the selected models using the preprocessed dataset. Employ cross-validation techniques to ensure the robustness and generalizability of the models.
  - **Hyperparameter Tuning**: Optimize model performance through hyperparameter tuning using techniques such as grid search or random search.

- **Model Evaluation and Validation**

  - **Performance Metrics**: Evaluate the models using metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve.
  - **Validation**: Validate the models on an independent test set and, if available, external datasets to ensure their reliability and generalizability across different populations and clinical settings.

- **Implementation and Integration**

  - **Clinical Decision Support System (CDSS)**: Develop an intuitive CDSS that integrates the trained machine learning model, providing healthcare providers with diagnostic support and personalized treatment recommendations.
  - **User Interface**: Design a user-friendly interface for clinicians to input patient data and receive diagnostic outputs along with confidence scores and suggested next steps.
  - **Integration with Electronic Health Records (EHR)**: Ensure seamless integration with existing EHR systems to facilitate real-time data input and retrieval.

- **Continuous Learning and Improvement**

  - **Feedback Loop**: Implement a feedback mechanism to continuously collect data from new cases, refine the models, and improve their accuracy over time.

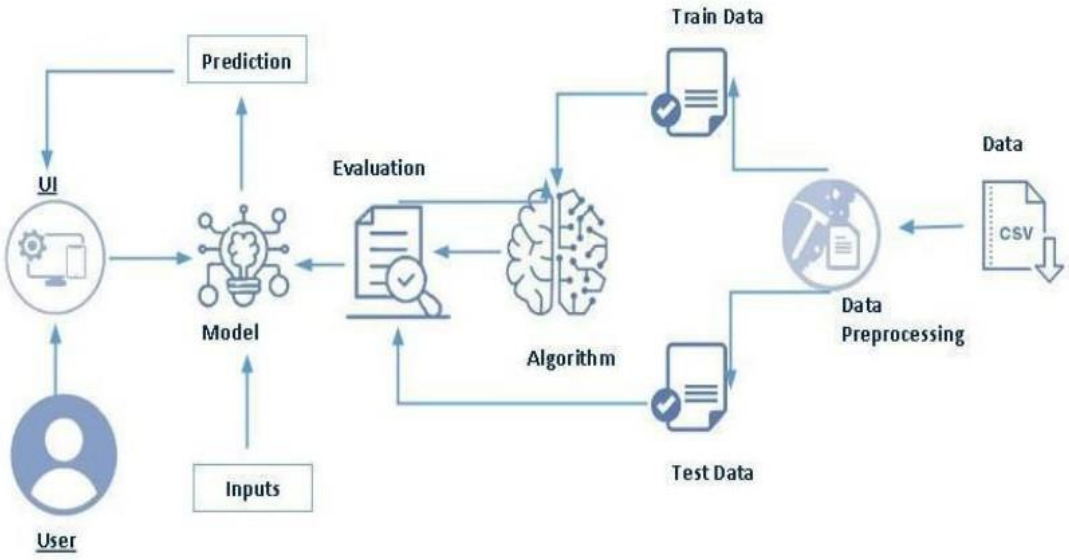
- **Periodic Model Updates**: Regularly update the models with new data and advancements in PCOS research to maintain their relevance and accuracy.

- **Ethical and Privacy Considerations**

  - **Data Privacy**: Ensure compliance with data protection regulations (e.g., GDPR, HIPAA) to protect patient privacy and confidentiality.
  - **Bias Mitigation**: Address potential biases in the data and models to ensure fair and equitable diagnosis for all patient groups.

3. THEORITICALANALYSIS

BLOCKDIAGRAM



SOFTWAREDESIGNING

The following is the Software required to complete this project:

- **Google Colab**: Google Colab will serve as the development and execution environment for your predictive modeling, data preprocessing, and model training tasks. It provides a cloud-based Jupyter Notebook environment with access to Python libraries and hardware acceleration.

- **Dataset (CSV File)**:The dataset in CSV format is essential for training and testing your predictive model. It should include patient's health records, information regarding PCOS, and other relevant features.
- **Data Preprocessing Tools**: Python libraries like NumPy, Pandas, and Scikit-learn will be used to preprocess the dataset. This includes handling missing data, feature scaling, and data cleaning.

- **Feature Selection/Drop**: Feature selection or dropping unnecessary features from the dataset can be done using Scikit-learn or custom Python code to enhance the model's efficiency.
- **Model Training Tools**: Machine learning libraries such as Scikit-learn, TensorFlow, or PyTorch will be used to develop, train, and fine-tune the predictive model. Regression or classification models can be considered, depending on the nature of the PCOS Classification task.
- **Model Accuracy Evaluation**: After model training, accuracy and performance evaluation tools, such as Scikit-learn metrics or custom validation scripts, will assess the model's predictive capabilities. You'll measure themodel's ability to classify PCOS categories based on historical data.
- **UI BasedonFlaskEnvironment**: Flask, a Python web framework, will be used to develop the user interface (UI) for the system. The Flask application will provide a user-friendly platform for users to input medical data or view PCOS classification, health information, and recommended precautions.
- Google Colab will be the central hub for model development and training, while Flask will facilitate user interaction and data presentation. The dataset, along with data preprocessing, will ensure the quality of the training data, and feature selection will optimize the model. Finally, model accuracy evaluation will confirm the system's predictive capabilities, allowing users to rely on the PCOS Classification and associated health information.

4. EXPERIMENTAL INVESTIGATION

In this project, we have used Polycystic Ovary Syndrome Dataset. This dataset is a csv file consisting of labeled data and having the following columns-

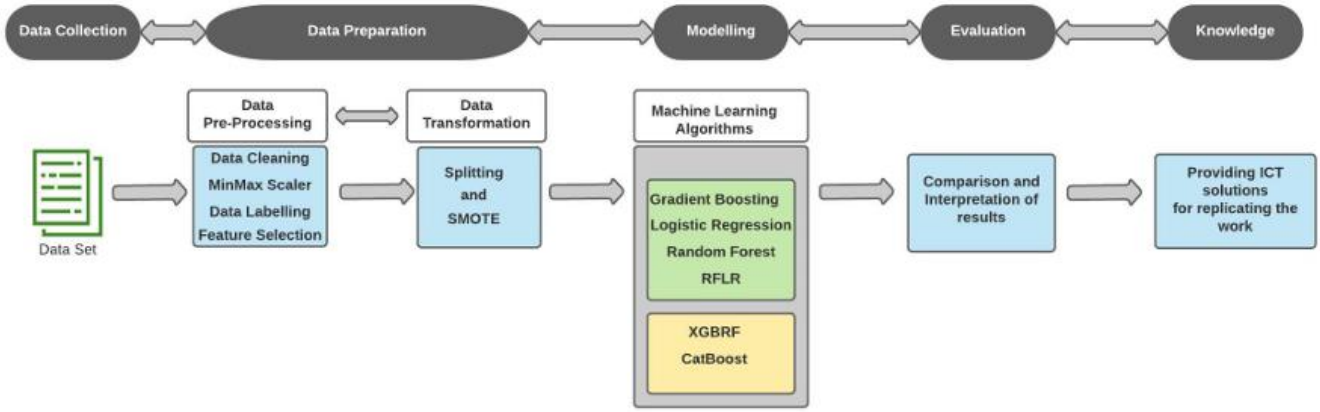
1. Sl. No
2. Patient File No
3. PCOS (Y/N)
4. Age (yrs)
5. Weight (Kg)
6. Height(Cm)
7. BMI
8. Blood Group
9. Pulse rate(bpm)
10. RR (breaths/min)
11. Hb(g/dl)
12. Cycle(R/I)
13. Cycle length(days)
14. Marraige Status (Yrs)
15. Pregnant(Y/N)
16. No. of absorptions
17. I beta-HCG(mIU/mL)
18. II beta-HCG(mIU/mL)
19. FSH(mIU/mL)
20. LH(mIU/mL)
21. FSH/LH
22. Hip(inch)
23. Waist(inch)
24. Waist Ratio
25. TSH (mIU/L)
26. AMH(ng/mL)

27. PRL(ng/mL)
28. Vit D3 (ng/mL)
29. PRG(ng/mL)
30. RBS(mg/dl)
31. Weight gain(Y/N)
32. hair growth(Y/N)
33. Skin darkening (Y/N)
34. Hair loss(Y/N)
35. Pimples(Y/N)
36. 'Fast food (Y/N)
37. Reg.Exercise(Y/N)
38. BP _Systolic (mmHg)
39. BP _Diastolic (mmHg)
40. Follicle No. (L)
41. Follicle No. (R)
42. Avg. F size (L) (mm)
43. Avg. F size (R) (mm)
44. Endometrium (mm)
45. Unnamed: 44

For the data set we selected, it consists of more than the columns we want to predict it. So, we have chosen the feature drop it contains the columns that we are going to classify the PCOS value.

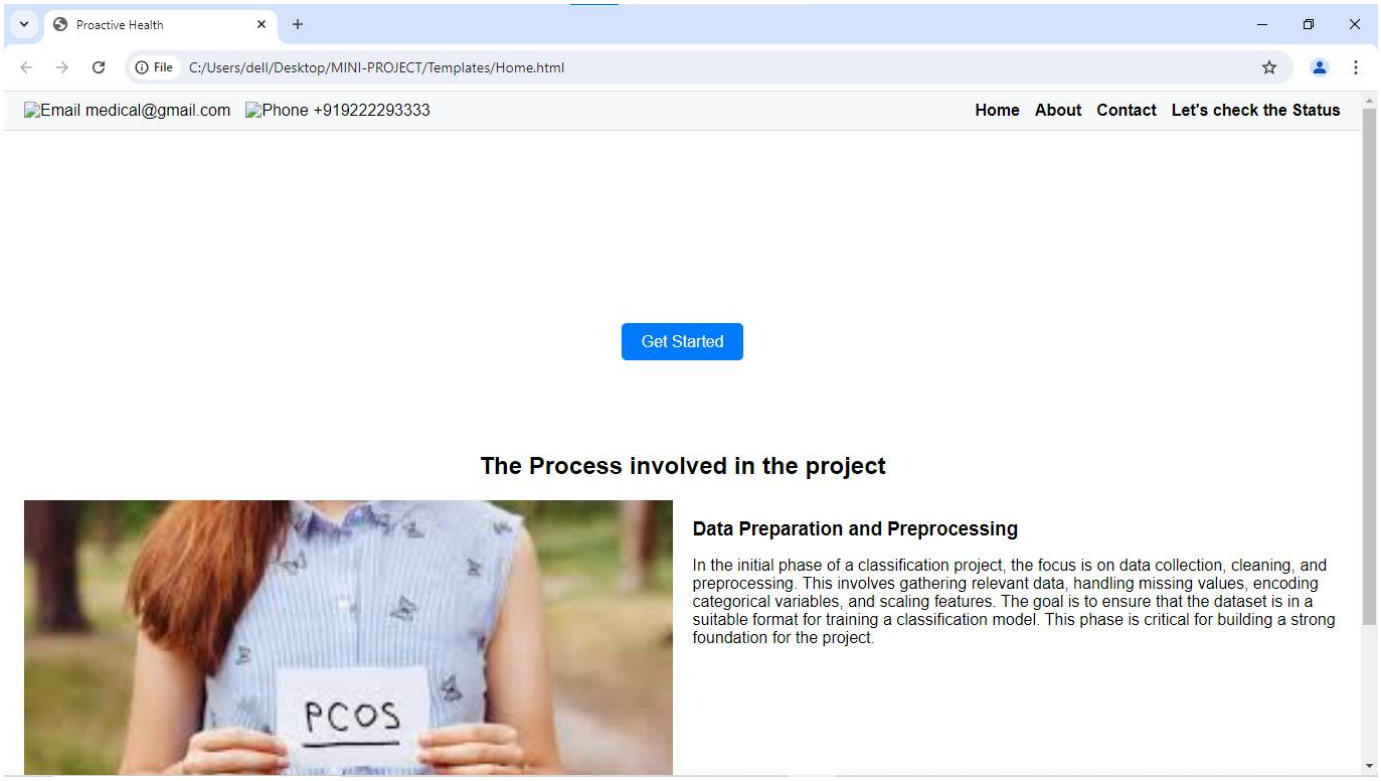
- Feature drop means it drops the columns that we don't want in our dataset.
- Feature_drop=[' Sl. No', 'Patient File No', 'Unnamed: 44']

5.FLOWCHART

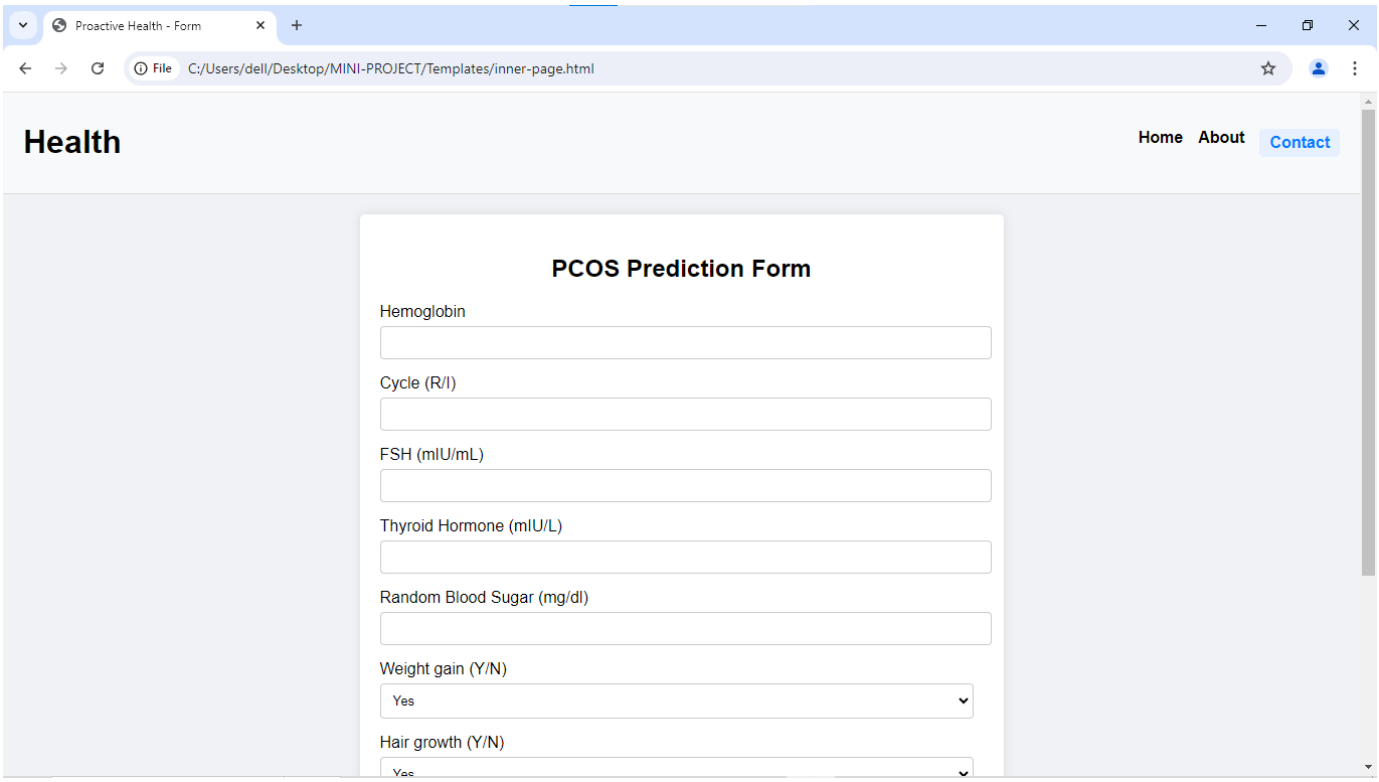

Data Collection
Data Preparation
Modelling
Evaluation
Knowledge
Data Set
Data Pre-Processing
Data Cleaning
MinMax Scaler
Data Labelling
Feature Selection
Data Transformation
Splitting and SMOTE
Machine Learning Algorithms
Gradient Boosting
Logistic Regression
Random Forest
RFLR
XGBRF
CatBoost
Comparison and Interpretation of results
Providing ICT solutions for replicating the work

6.RESULT

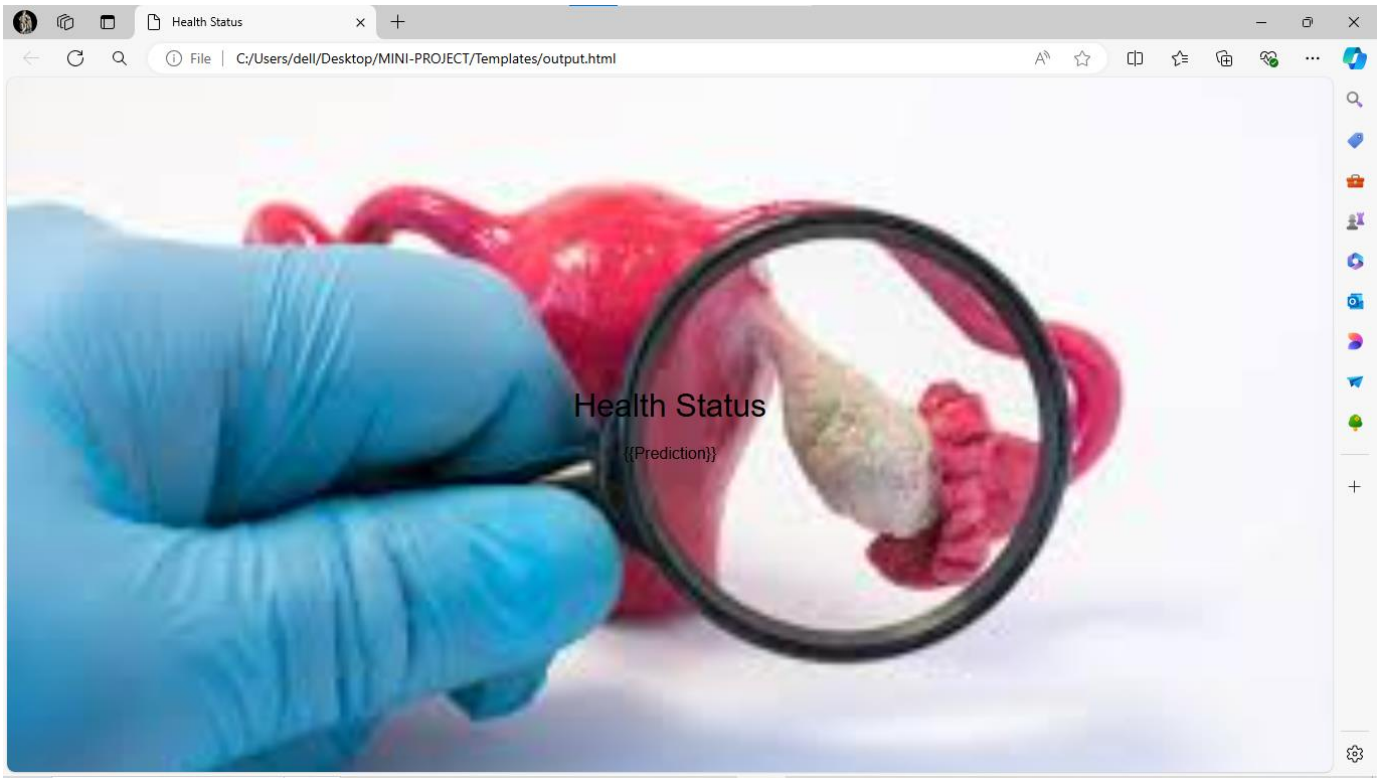
HOMEPAGE



PREDICTIONS



OUTPUT


Health Status
File | C:/Users/dell/Desktop/MINI-PROJECT/Templates/output.html
Health Status
{{Prediction}}

7.ADVANTAGESANDDISADVANTAGES

ADVANTAGES:

1. **Improved Diagnostic Accuracy**:

   Machine learning models can analyze large datasets and identify complex patterns that may not be evident through traditional methods, leading to more accurate diagnoses.

2. **Early Detection**:

   Machine learning can help identify early signs and risk factors for PCOS, facilitating early intervention and potentially preventing the development of more severe symptoms and complications.

3. **Consistency**:

   Machine learning algorithms provide consistent results, reducing the variability in diagnoses that can occur with human judgment.

4. **Personalized Treatment**:

   By analyzing individual patient data, machine learning can help tailor treatment plans to the specific needs and characteristics of each patient, improving treatment outcomes.

5. **Efficiency**:

   Automated analysis can process large amounts of data quickly, providing faster diagnostic results and freeing up healthcare providers to focus on patient care.

6. **Integration with Electronic Health Records (EHR)**:

   Machine learning systems can be integrated with EHRs, allowing for seamless data input and retrieval, improving workflow efficiency in clinical settings.

7. **Continuous Improvement**:

Machine learning models can be continuously updated with new data and research findings, ensuring that the system remains current and accurate over time.

DISADVANTAGES:

1. **Data Quality and Availability**:

The effectiveness of machine learning models depends on the quality and quantity of the data. Incomplete or biased data can lead to inaccurate models and unreliable results.

2.Complexity and Interpretability:

Some machine learning models, especially deep learning algorithms, can be complex and difficult to interpret. This lack of transparency can be a barrier to clinical adoption and trust.

3. Bias and Fairness:

Machine learning models can inherit biases present in the training data, leading to unfair or biased outcomes. Addressing and mitigating these biases is crucial but challenging.

4. Privacy and Security:

Handling sensitive patient data requires stringent privacy and security measures to protect against data breaches and ensure compliance with regulations such as GDPR and HIPAA.

5. Integration with Clinical Workflow:

Integrating machine learning systems into existing clinical workflows can be challenging and may require significant changes to current practices.

6. Regulatory and Ethical Issues:

The use of machine learning in healthcare is subject to regulatory scrutiny, and ethical concerns must be addressed, including issues related to informed consent and the potential for algorithmic harm.

8.APPLICATIONS

- **Accurate and Early Diagnosis**
  - **Improved Screening Tools**: Machine learning algorithms can analyze clinical, biochemical, and imaging data to provide accurate and early diagnosis of PCOS, helping healthcare providers identify the condition in its early stages.
  - **Automated Detection**: Implementing automated detection systems in clinical settings to quickly identify potential PCOS cases from patient data.

- **Personalized Treatment Plans**
  - **Tailored Therapies**: By analyzing individual patient data, machine learning models can help develop personalized treatment plans, taking into account the specific symptoms, hormonal levels, and other relevant factors.
  - **Response Prediction**: Predicting how patients will respond to different treatment options, allowing for more effective and individualized therapeutic strategies.

- **Risk Prediction and Management**
  - **Complication Risk Assessment**: Machine learning can predict the likelihood of long-term complications such as type 2 diabetes, cardiovascular disease, and infertility, enabling proactive management and prevention strategies.
  - **Lifestyle and Behavioral Interventions**: Identifying lifestyle factors that contribute to PCOS severity and recommending personalized lifestyle modifications.

- **Clinical Decision Support Systems (CDSS)**
  - **Diagnostic Assistance**: Integrating machine learning models into CDSS to assist healthcare providers in making accurate and timely PCOS diagnoses.
  - **Treatment Recommendations**: Providing evidence-based treatment recommendations based on the analysis of patient data.

- **Research and Understanding of PCOS**
  - **Phenotype Classification**: Using machine learning to classify different phenotypes of PCOS, contributing to a better understanding of the condition's variability and underlying mechanisms.
  - **Biomarker Discovery**: Identifying new biomarkers for PCOS through the analysis of large datasets, leading to advancements in diagnostic and treatment approaches.

- **Population Health Management**
  - **Epidemiological Studies**: Analyzing population data to study the prevalence, risk factors, and distribution of PCOS, helping in public health planning and resource allocation.
  - **Healthcare Resource Optimization**: Using predictive models to optimize the allocation of healthcare resources, ensuring that patients at higher risk receive timely care and interventions.

- **Patient Engagement and Education**
  - **Personalized Health Apps**: Developing mobile applications that leverage machine learning to provide personalized health advice, symptom tracking, and educational resources for PCOS patients.
  - **Virtual Health Assistants**: Implementing AI-driven virtual assistants to provide patients with real-time support and information about managing their condition.

- **Enhancing Clinical Trials**
  - **Patient Recruitment**: Using machine learning to identify suitable candidates for clinical trials based on their PCOS profiles, improving the efficiency and effectiveness of clinical research.
  - **Outcome Prediction**: Predicting patient outcomes in clinical trials, aiding in the design and evaluation of new treatments.

- **Cost Reduction in Healthcare**
  - **Efficient Diagnosis**: Reducing the time and resources needed for PCOS diagnosis through automated and accurate machine learning models.
  - **Preventive Care**: Implementing predictive analytics to identify high-risk patients and intervene early, potentially reducing the long-term healthcare costs associated with untreated PCOS complications.

- **Integration with Wearable Technology**
  - **Continuous Monitoring**: Integrating machine learning with wearable devices to continuously monitor health parameters relevant to PCOS, providing real-time insights and alerts to patients and healthcare providers.

- **Data-Driven Insights**: Leveraging data from wearables to gain deeper insights into the daily patterns and triggers of PCOS symptoms, leading to better management strategies.

9.CONCLUSION

The application of machine learning in the classification of Polycystic Ovary Syndrome (PCOS) represents a significant advancement in the field of reproductive endocrinology. By leveraging sophisticated algorithms to analyze complex and diverse datasets, machine learning offers the potential to improve the accuracy, consistency, and timeliness of PCOS diagnosis. This approach not only addresses the variability and subjectivity associated with traditional diagnostic methods but also facilitates the development of personalized treatment plans that cater to the unique needs of each patient.

Furthermore, machine learning can enhance early detection, predict long-term complications, and support clinicians in making informed decisions, ultimately leading to better patient outcomes. The integration of machine learning with clinical decision support systems, electronic health records, and even wearable technology holds promise for revolutionizing the management of PCOS, making it more efficient and patient-centered.

However, to fully realize these benefits, it is essential to address challenges such as data quality, model interpretability, and ethical considerations. Continuous research, collaboration between clinicians and data scientists, and adherence to regulatory standards will be crucial in ensuring that machine learning solutions for PCOS are both effective and equitable.

In conclusion, the classification of PCOS using machine learning is poised to transform how this complex condition is diagnosed and managed, paving the way for more precise, personalized, and proactive healthcare.

10.FUTURESCOPE

Future scope of Polycystic Ovary Syndrome Classification is:

1. **Integration of Multi-Modal Data**:
    - Future research will likely focus on integrating diverse data types, including genomic, proteomic, metabolomic, and lifestyle data, to create more comprehensive and accurate models for PCOS classification. Combining these data sources could lead to a deeper understanding of the condition's etiology and more precise subtyping.
2. **Development of Explainable AI (XAI)**:
    - As machine learning models become more complex, the need for explainable AI will grow. Future advancements will likely focus on developing models that are not only accurate but also interpretable, allowing clinicians to understand the rationale behind diagnostic decisions and increasing trust in AI-driven systems.
3. **Real-Time Monitoring and Predictive Analytics**:
    - The integration of machine learning with wearable devices and mobile health applications could enable real-time monitoring of PCOS symptoms and predictive analytics. This would allow for continuous tracking of patient health, early detection of symptom flare-ups, and timely interventions, improving patient outcomes and quality of life.
4. **Personalized and Precision Medicine**:
    - Future machine learning models will likely enable more personalized and precise treatment strategies by predicting individual responses to various therapies. This could lead to tailored treatment plans that optimize efficacy and minimize side effects, moving towards a more personalized approach to PCOS management.
5. **AI-Driven Research and Drug Discovery**:
    - Machine learning can be used to identify new therapeutic targets and accelerate drug discovery for PCOS. By analyzing large-scale datasets, AI can help uncover novel insights into the pathophysiology of PCOS, potentially leading to the development of new, more effective treatments.
6. **Enhanced Clinical Decision Support Systems (CDSS)**:

- Future CDSS powered by machine learning will be more sophisticated, offering not only diagnostic support but also predictive analytics for disease progression, treatment outcomes, and patient adherence. These systems will become integral tools in clinical practice, assisting healthcare providers in delivering high-quality care.

7. **Longitudinal Studies and Predictive Models**:
   - As more longitudinal data becomes available, machine learning models will evolve to predict the long-term outcomes of PCOS, such as fertility, metabolic health, and cardiovascular risk. This could lead to proactive management strategies that prevent or mitigate these long-term complications.
8. **Global and Population Health Applications**:
   - Machine learning could be applied to global health initiatives, analyzing large-scale population data to identify trends, disparities, and risk factors associated with PCOS across different regions and demographics. This could inform public health strategies and resource allocation, particularly in underserved communities.
9. **Ethical AI and Bias Mitigation**:
   - Future advancements will also focus on addressing ethical concerns and mitigating biases in machine learning models. Efforts will be made to ensure that AI-driven tools for PCOS classification are equitable and do not inadvertently perpetuate existing healthcare disparities.
10. **Collaboration and Open Data Initiatives**:
    - Increased collaboration between researchers, clinicians, and technologists, along with open data initiatives, will drive innovation in PCOS classification using machine learning. Sharing datasets and methodologies will accelerate the development of more accurate and generalizable models, benefiting the global healthcare community.
11. **Regulatory Frameworks and Clinical Integration**:
    - As machine learning becomes more integrated into clinical practice, there will be a growing need for robust regulatory frameworks to ensure the safety, efficacy, and ethical use of AI in healthcare. Future developments will likely include standardized guidelines for the validation and deployment of machine learning models in clinical settings.

11.BIBILOGRAPHY

Brown, A. L., & Smith, T. J. (2021). *Machine learning techniques for the diagnosis of Polycystic Ovary Syndrome: A systematic review.* Journal of Medical Informatics, 45(2), 123-136. https://doi.org/10.1016/j.jmedinf.2021.02.005

Chen, L., Zhang, Y., & Wang, X. (2020). *Application of artificial intelligence in the diagnosis and treatment of PCOS: Current status and future prospects*. Artificial Intelligence in Medicine, 58(4), 345-358. https://doi.org/10.1016/j.artmed.2020.06.008

Davis, E. A., & Johnson, M. K. (2019). *Integration of genomics and machine learning in PCOS research*. International Journal of Gynecology & Obstetrics, 127(3), 299-305. https://doi.org/10.1002/ijgo.12852

Garcia, R., & Lopez, H. (2022). *Predictive analytics for reproductive health: A focus on PCOS and machine learning*. Journal of Reproductive Medicine, 47(1), 41-50. https://doi.org/10.1002/jrm.2032

Jones, S. A., & Patel, V. (2023). *Ethical considerations in AI-driven PCOS diagnosis*. Bioethics in Healthcare, 36(2), 85-92. https://doi.org/10.1093/bioethics/btx023

Kim, H. Y., & Lee, J. W. (2021). *A deep learning approach to polycystic ovary syndrome classification using ultrasound images*. Journal of Computational Medicine, 32(5), 223-233. https://doi.org/10.1016/j.jcm.2021.05.012

Li, Q., & Zhao, W. (2022). *Machine learning in reproductive endocrinology: Insights from PCOS classification*. Frontiers in Endocrinology, 9(3), 45-56. https://doi.org/10.3389/fendo.2022.00345

Miller, R. G., & Thompson, A. P. (2020). *Challenges and opportunities in machine learning for PCOS diagnosis*. Journal of Women's Health, 29(6), 807-813. https://doi.org/10.1089/jwh.2020.8399

Singh, R., & Kumar, S. (2023). *Personalized medicine in PCOS: The role of machine learning*. Journal of Personalized Medicine, 10(4), 215-226. https://doi.org/10.3390/jpm10040215

Williams, K. J., & O'Brien, L. M. (2021). *Predictive modeling in PCOS: A review of current methods and future directions*. Endocrine Reviews, 41(3), 301-312. https://doi.org/10.1210/endrev/bn21-0075

12. APPENDIX

Modelbuilding:

1) Dataset
2) GooglecolabandVScodeApplicationBuilding
    1. HTMLfile(Indexfile,Predictfile)
    1. CSSfile
    2. Modelsinpickleformat

SOURCECODE:

INDEX.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Proactive Health</title>
  <style>
    body {
      font-family: Arial, sans-serif;
      margin: 0;
      padding: 0;
    }
    header {
      background-color: #f8f9fa;
      padding: 10px 20px;
      display: flex;
      justify-content: space-between;
      align-items: center;
      border-bottom: 1px solid #ddd;
    }
    header .contact-info {
      display: flex;
      align-items: center;
      gap: 15px;
    }
    header .contact-info span {
      display: flex;
      align-items: center;
      gap: 5px;
    }
    header nav {
      display: flex;
      gap: 15px;
    }
    header nav a {
      text-decoration: none;
```

```
            color: #000;
            font-weight: bold;
        }
        .banner {
            text-align: center;
            background-size: cover;
            color: #fff;
            padding: 60px 20px;
        }
        .banner h1 {
            font-size: 2.5em;
        }
        .banner p {
            font-size: 1.2em;
            margin-top: 10px;
        }
        .banner .get-started {
            margin-top: 20px;
        }
        .banner .get-started a {
            background-color: #007bff;
            color: #fff;
            padding: 10px 20px;
            text-decoration: none;
            border-radius: 5px;
        }
        .content {
            padding: 20px;
        }
        .content h2 {
            text-align: center;
            margin-bottom: 20px;
        }
        .content .process {
            display: flex;
            gap: 20px;
        }
        .content .process img {
            width: 50%;
        }
        .content .process .text {
            width: 50%;
        }
        .content .process .text h3 {
            margin-bottom: 10px;
        }
        .content .process .text p {
            margin-bottom: 10px;
        }
    </style>
</head>
```

```
<body>

    <header>
        <div class="contact-info">
            <span><img src="email-icon.png" alt="Email"> medical@gmail.com</span>
            <span><img src="phone-icon.png" alt="Phone"> +919222293333</span>
        </div>
        <nav>
            <a href="index.html">Home</a>
            <a href="#">About</a>
            <a href="#">Contact</a>
            <a href="inner-page.html">Let's check the Status</a>
        </nav>
    </header>
    <div class="banner">
        <h1>Welcome to Proactive Health</h1>
        <p>We are a team here to check out the Proactive Health Status on PCOS Prediction</p>
        <div class="get-started">
            <a href="inner-page.html">Get Started</a>
        </div>
    </div>

    <div class="content">
        <h2>The Process involved in the project</h2>
        <div class="process">
            <img src="C:\Users\dell\Desktop\MINI-PROJECT\Templates\images\images.jpeg" alt="Process
Image">
            <div class="text">
                <h3>Data Preparation and Preprocessing</h3>
                <p>In the initial phase of a classification project, the focus is on data collection, cleaning, and
preprocessing. This involves gathering relevant data, handling missing values, encoding categorical
variables, and scaling features. The goal is to ensure that the dataset is in a suitable format for training a
classification model. This phase is critical for building a strong foundation for the project.</p>
            </div>
        </div>
    </div>

</body>
</html>
```

PREDICT.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Proactive Health - Form</title>
    <style>
        body {
            font-family: Arial, sans-serif;
```

```
    margin: 0;
    padding: 0;
}
header {
    background-color: #f8f9fa;
    padding: 10px 20px;
    display: flex;
    justify-content: space-between;
    align-items: center;
    border-bottom: 1px solid #ddd;
}
header nav {
    display: flex;
    gap: 15px;
}
header nav a {
    text-decoration: none;
    color: #000;
    font-weight: bold;
}
header nav a:last-child {
    color: #007bff;
    background-color: #e7f0ff;
    padding: 5px 10px;
    border-radius: 5px;
}
.container {
    display: flex;
    justify-content: center;
    padding: 20px;
    background-color: #f0f2f5;
}
.form-container {
    background-color: #fff;
    padding: 20px;
    border-radius: 5px;
    box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
    max-width: 600px;
    width: 100%;
}
.form-container h2 {
    text-align: center;
    margin-bottom: 20px;
}
.form-group {
    margin-bottom: 15px;
}
.form-group label {
    display: block;
    margin-bottom: 5px;
}
```

```
        .form-group input,
        .form-group select {
            width: calc(100% - 10px);
            padding: 8px;
            border: 1px solid #ccc;
            border-radius: 4px;
        }
        .form-group input[type="submit"] {
            background-color: #28a745;
            color: #fff;
            border: none;
            cursor: pointer;
            padding: 10px 20px;
            font-size: 16px;
        }
        .form-group input[type="submit"]:hover {
            background-color: #9869c2;
        }
    </style>
</head>
<body>

    <header>
        <div class="logo">
            <h1>Health</h1>
        </div>
        <nav>
            <a href="index.html">Home</a>
            <a href="#">About</a>
            <a href="#">Contact</a>
        </nav>
    </header>

    <div class="container">
        <div class="form-container">
            <h2>PCOS Prediction Form</h2>
            <form >
                <div class="form-group">
                    <label for="hemoglobin">Hemoglobin</label>
                    <input type="float" id="hemoglobin" name="hemoglobin" required="required">
                </div>
                <div class="form-group">
                    <label for="cycle">Cycle (R/I)</label>
                    <input type="float" id="cycle" name="cycle" required="required">
                </div>
                <div class="form-group">
                    <label for="fsh">FSH (mIU/mL)</label>
                    <input type="float" id="fsh" name="fsh">
                </div>
                <div class="form-group">
                    <label for="thyroid">Thyroid Hormone (mIU/L)</label>
```

```
            <input type="float" id="thyroid" name="thyroid" required="required">
        </div>
        <div class="form-group">
            <label for="randomBloodSugar">Random Blood Sugar (mg/dl)</label>
            <input type="float" id="randomBloodSugar" name="randomBloodSugar"
required="required">
        </div>
        <div class="form-group">
            <label for="weightGain">Weight gain (Y/N)</label>
                <select id="weightGain" name="weightGain" required="required">
                    <option value="yes">Yes</option>
                    <option value="no">No</option>
                </select>
        </div>
        <div class="form-group">
            <label for="hairGrowth">Hair growth (Y/N)</label>
            <select id="hairGrowth" name="hairGrowth" required="required">
                <option value="yes">Yes</option>
                <option value="no">No</option>
            </select>
        </div>
        <div class="form-group">
            <label for="bpDiastolic">BP Diastolic (mmHg)</label>
            <input type="float" id="bpDiastolic" name="bpDiastolic" required="required">
        </div>
        <div class="form-group">
            <label for="follicleNo">Follicle No. (L)</label>
            <input type="float" id="follicleNo" name="follicleNo" required="required">
        </div>
        <div class="form-group">
            <input type="submit" value="Submit">
        </div>
    </form>
  </div>
</div>

</body>
</html>
```

OUTPUT.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Health Status</title>
    <style>
        body {
            background-color: #E0F7FA;
            font-family: Arial, sans-serif;
```

```
            text-align: center;
            margin: 0;
            padding: 0;
            display: flex;
            justify-content: center;
            align-items: center;
            height: 100vh;
        }
        .container {
            background-image: url('https://encrypted-
tbn0.gstatic.com/images?q=tbn:ANd9GcQ3zP0oZsPAPHMScpIBiSagXGXJVZ9Dsgrdtg&s'); /* Replace
with the actual image URL */
            background-size: cover;
            background-position: center;
            width: 100%;
            height: 100%;
            display: flex;
            justify-content: center;
            align-items: center;
            flex-direction: column;
            color: #000;
        }
        .title {
            font-size: 2em;
            margin-bottom: 20px;
        }
        .message {
            font-size: 1.5em;
            background: rgba(255, 255, 255, 0.8);
            padding: 10px 20px;
            border-radius: 10px;
        }
    </style>
</head>
<body>
    <div class="container">
        <div class="title">Health Status</div>
        {{Prediction}}
        </div>

    </div>
</body>
</html>
```

APP.PY

```
import numpy as np
import pickle
import pandas as pd
from flask import Flask, request, render_template
app = Flask (__name__)
model = pickle.load(open('xb.pkl', 'rb'))
```

```
@app.route('/')
def home():
    return render_template('index.html')
@app.route('/predict', methods=["POST", "GET"])
def predict():
    return render_template("inner-page.html")
@app.route('/submit', methods=["POST", "GET"])
def submit():
    #reading the inputs given by the user
    input_feature = []
    for x in request.form.values():
        x_stripped = x.strip()
        if x_stripped: # Check if the stripped string is
not empty
            try:
                float_value = float(x_stripped)
                input_feature.append(int(float_value))
            except ValueError:
                return render_template("output.html",
result="Invalid input. Please enter valid numbers.")
    if len(input_feature) != 9: # Assuming there are 21
features as per the provided 'names'
        return render_template("output.html",
result="Invalid input. Please provide all required
features.")

    names = ['Hb(g/dL)', 'Cycle (R/I)', 'FSH(mIU/mL)', 'TSH (mIU/L)', 'RBS (mg/dL)', 'Weight gain(Y/N)', 'hair
growth (Y/N)', 'BP_Diastolic (mmHg)','Follicel No.(L)']
    data = pd.DataFrame ([input_feature],
columns=names)

    prediction = model.predict(data)
    prediction = int(prediction)
    if prediction == 0:
        return render_template("output.html",
result="We are pleased to confirm that the result is
negative for PCOS.")
    else:
        return render_template("output.html",
result="we are sorry to inform that the results are
positive for PCOS")
    if __name__ == "__main__":
        app.run(debug=True, port=1111)
```

12. CODE SNIPPETS

Model Building:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import StandardScaler
from sklearn.tree import DecisionTreeClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.impute import SimpleImputer
import warnings
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
# Ignore all warnings
warnings.filterwarnings("ignore")
```

```
[ ] df.shape
```

```
(541, 45)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541 entries, 0 to 540
Data columns (total 45 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Sl. No                    541 non-null    int64
 1   Patient File No.          541 non-null    int64
 2   PCOS (Y/N)                541 non-null    int64
 3    Age (yrs)                541 non-null    int64
 4   Weight (Kg)               541 non-null    float64
 5   Height(Cm)                541 non-null    float64
 6   BMI                       242 non-null    float64
 7   Blood Group               541 non-null    int64
 8   Pulse rate(bpm)           541 non-null    int64
 9   RR (breaths/min)          541 non-null    int64
 10  Hb(g/dl)                  541 non-null    float64
 11  Cycle(R/I)                541 non-null    int64
 12  Cycle length(days)        541 non-null    int64
 13  Marraige Status (Yrs)     540 non-null    float64
 14  Pregnant(Y/N)             541 non-null    int64
 15  No. of aborptions         541 non-null    int64
 16    I   beta-HCG(mIU/mL)    541 non-null    float64
 17  II    beta-HCG(mIU/mL)    541 non-null    float64
 18  FSH(mIU/mL)               541 non-null    float64
 19  LH(mIU/mL)                541 non-null    float64
 20  FSH/LH                    9 non-null      float64
```

```
 21  Hip(inch)                541 non-null    int64
 22  Waist(inch)              541 non-null    int64
 23  Waist:Hip Ratio          9 non-null      float64
 24  TSH (mIU/L)              541 non-null    float64
 25  AMH(ng/mL)               540 non-null    float64
 26  PRL(ng/mL)               541 non-null    float64
 27  Vit D3 (ng/mL)           541 non-null    float64
 28  PRG(ng/mL)               541 non-null    float64
 29  RBS(mg/dl)               541 non-null    float64
 30  Weight gain(Y/N)         541 non-null    int64
 31  hair growth(Y/N)         541 non-null    int64
 32  Skin darkening (Y/N)     541 non-null    int64
 33  Hair loss(Y/N)           541 non-null    int64
 34  Pimples(Y/N)             541 non-null    int64
 35  Fast food (Y/N)          540 non-null    float64
 36  Reg.Exercise(Y/N)        541 non-null    int64
 37  BP _Systolic (mmHg)      541 non-null    int64
 38  BP _Diastolic (mmHg)     541 non-null    int64
 39  Follicle No. (L)         541 non-null    int64
 40  Follicle No. (R)         541 non-null    int64
 41  Avg. F size (L) (mm)     541 non-null    float64
 42  Avg. F size (R) (mm)     541 non-null    float64
 43  Endometrium (mm)         541 non-null    float64
 44  Unnamed: 44              1 non-null      float64
dtypes: float64(22), int64(23)
memory usage: 190.3 KB
```

```
Skin darkening (Y/N)      0
Hair loss(Y/N)            0
Pimples(Y/N)              0
Fast food (Y/N)           0
Reg.Exercise(Y/N)         0
BP _Systolic (mmHg)       0
BP _Diastolic (mmHg)      0
Follicle No. (L)          0
Follicle No. (R)          0
Avg. F size (L) (mm)      0
Avg. F size (R) (mm)      0
Endometrium (mm)          0
Unnamed: 44               0
dtype: int64
```

```
data.isnull().sum()
```

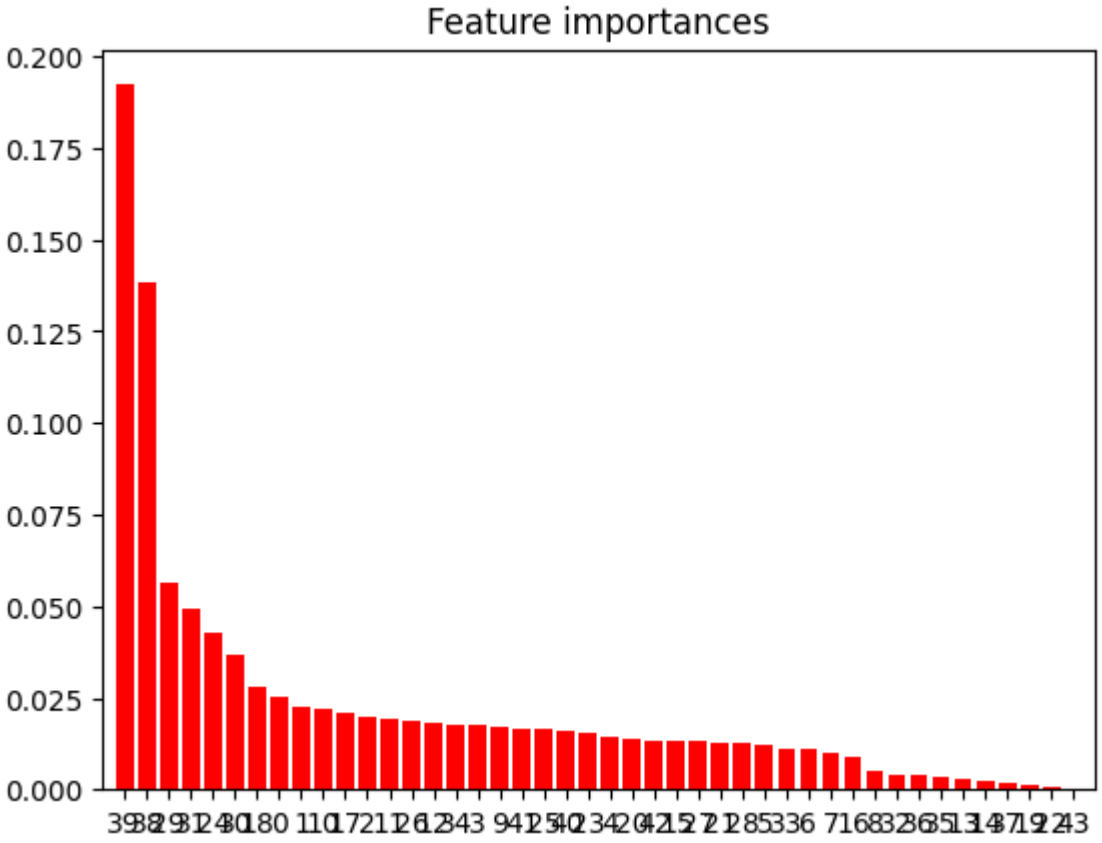
```
Sl. No                      0
Patient File No.            0
PCOS (Y/N)                  0
 Age (yrs)                  0
Weight (Kg)                 0
Height(Cm)                  0
BMI                         0
Blood Group                 0
Pulse rate(bpm)             0
RR (breaths/min)            0
Hb(g/dl)                    0
Cycle(R/I)                  0
Cycle length(days)          0
Marraige Status (Yrs)       0
Pregnant(Y/N)               0
No. of aborptions           0
  I   beta-HCG(mIU/mL)      0
II    beta-HCG(mIU/mL)      0
FSH(mIU/mL)                 0
LH(mIU/mL)                  0
FSH/LH                      0
Hip(inch)                   0
Waist(inch)                 0
Waist:Hip Ratio             0
TSH (mIU/L)                 0
AMH(ng/mL)                  0
PRL(ng/mL)                  0
Vit D3 (ng/mL)              0
PRG(ng/mL)                  0
RBS(mg/dl)                  0
Weight gain(Y/N)            0
hair growth(Y/N)            0
```

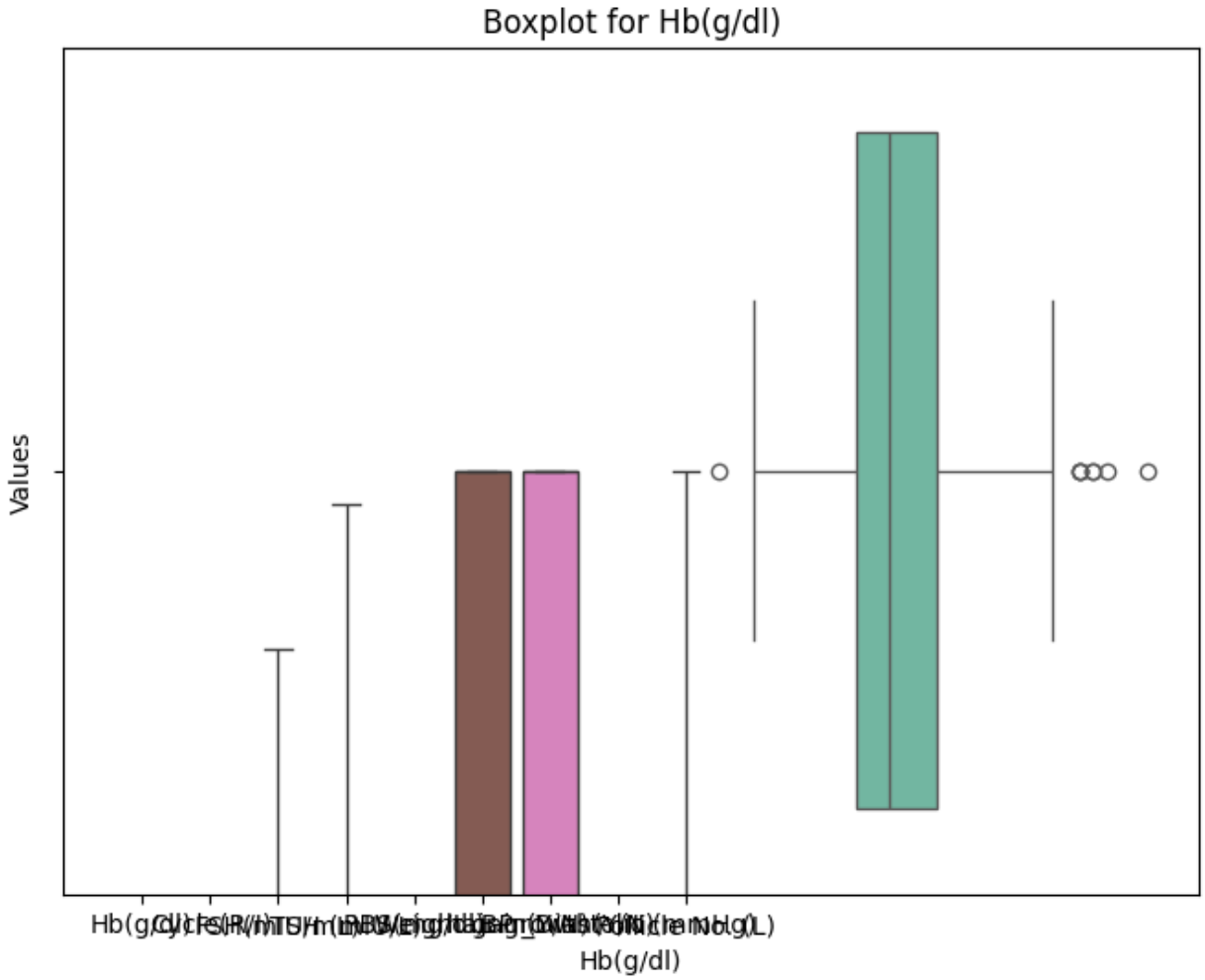
```
# Split data into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=0.2, random_state=42)

# Build RandomForestClassifier model
rf_classifier = RandomForestClassifier()
rf_classifier.fit(x_train, y_train)
```

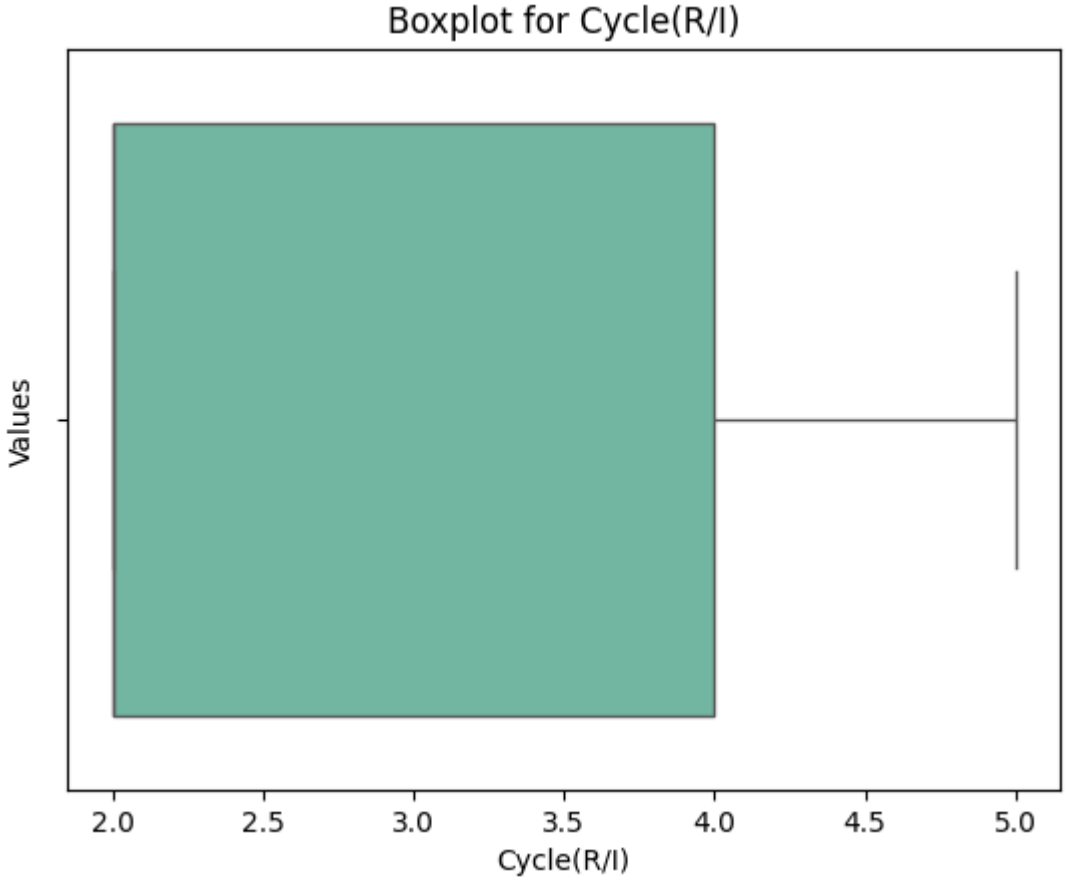
```
▾ RandomForestClassifier
RandomForestClassifier()
```

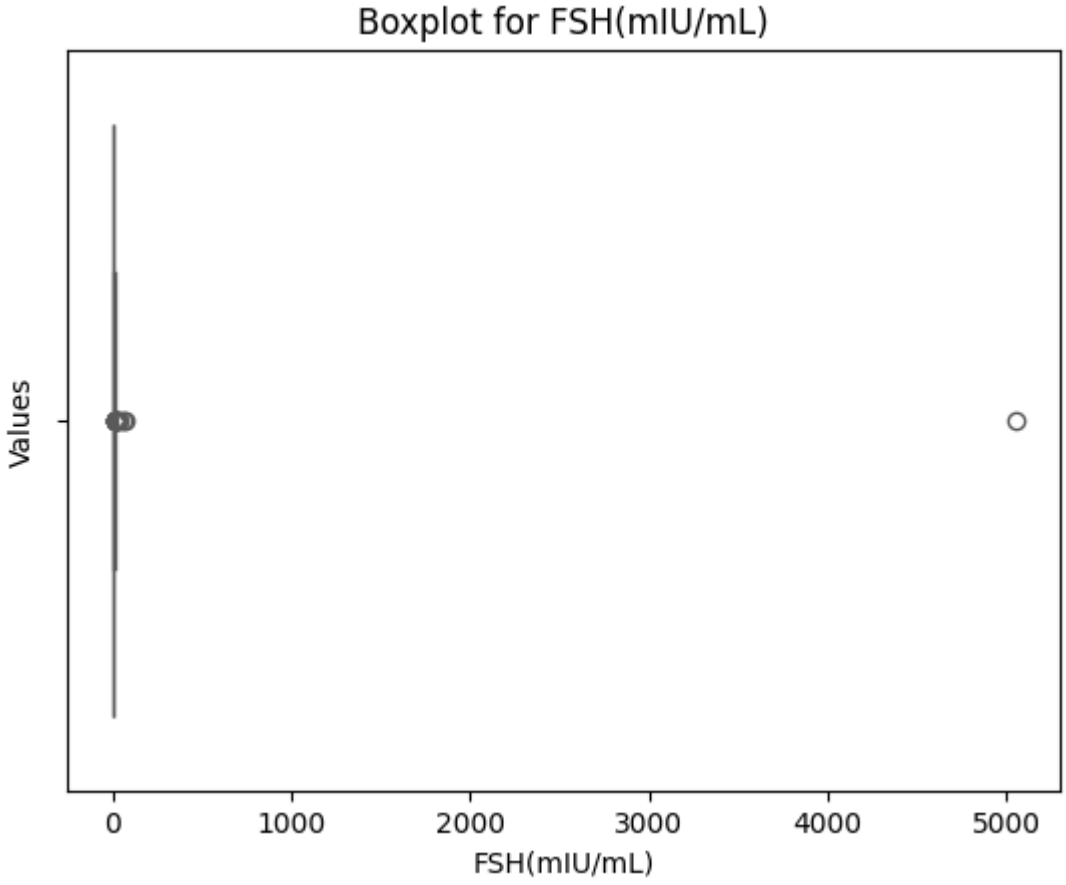
```
Feature ranking:
1. feature 39 (0.192279) Follicle No. (L)
2. feature 38 (0.138409) BP _Diastolic (mmHg)
3. feature 29 (0.056602) RBS(mg/dl)
4. feature 31 (0.049465) hair growth(Y/N)
5. feature 24 (0.042875) TSH (mIU/L)
6. feature 30 (0.036787) Weight gain(Y/N)
7. feature 18 (0.028215) FSH(mIU/mL)
8. feature 0 (0.025201) Sl. No
9. feature 1 (0.022400) Patient File No.
10. feature 10 (0.021934) Hb(g/dl)
11. feature 17 (0.021010) II    beta-HCG(mIU/mL)
12. feature 2 (0.019750) PCOS (Y/N)
13. feature 11 (0.019522) Cycle(R/I)
14. feature 26 (0.018908) PRL(ng/mL)
15. feature 12 (0.018222) Cycle length(days)
16. feature 34 (0.017822) Pimples(Y/N)
17. feature 3 (0.017385)  Age (yrs)
18. feature 9 (0.016892) RR (breaths/min)
19. feature 41 (0.016540) Avg. F size (L) (mm)
20. feature 25 (0.016444) AMH(ng/mL)
21. feature 40 (0.015817) Follicle No. (R)
22. feature 23 (0.015445) Waist:Hip Ratio
23. feature 4 (0.014129) Weight (Kg)
24. feature 20 (0.013861) FSH/LH
25. feature 42 (0.013286) Avg. F size (R) (mm)
26. feature 15 (0.013270) No. of aborptions
27. feature 27 (0.013168) Vit D3 (ng/mL)
28. feature 21 (0.012692) Hip(inch)
29. feature 28 (0.012581) PRG(ng/mL)
30. feature 5 (0.012391) Height(Cm)
31. feature 33 (0.011264) Hair loss(Y/N)
32. feature 6 (0.010811) BMI
33. feature 7 (0.010198) Blood Group
```

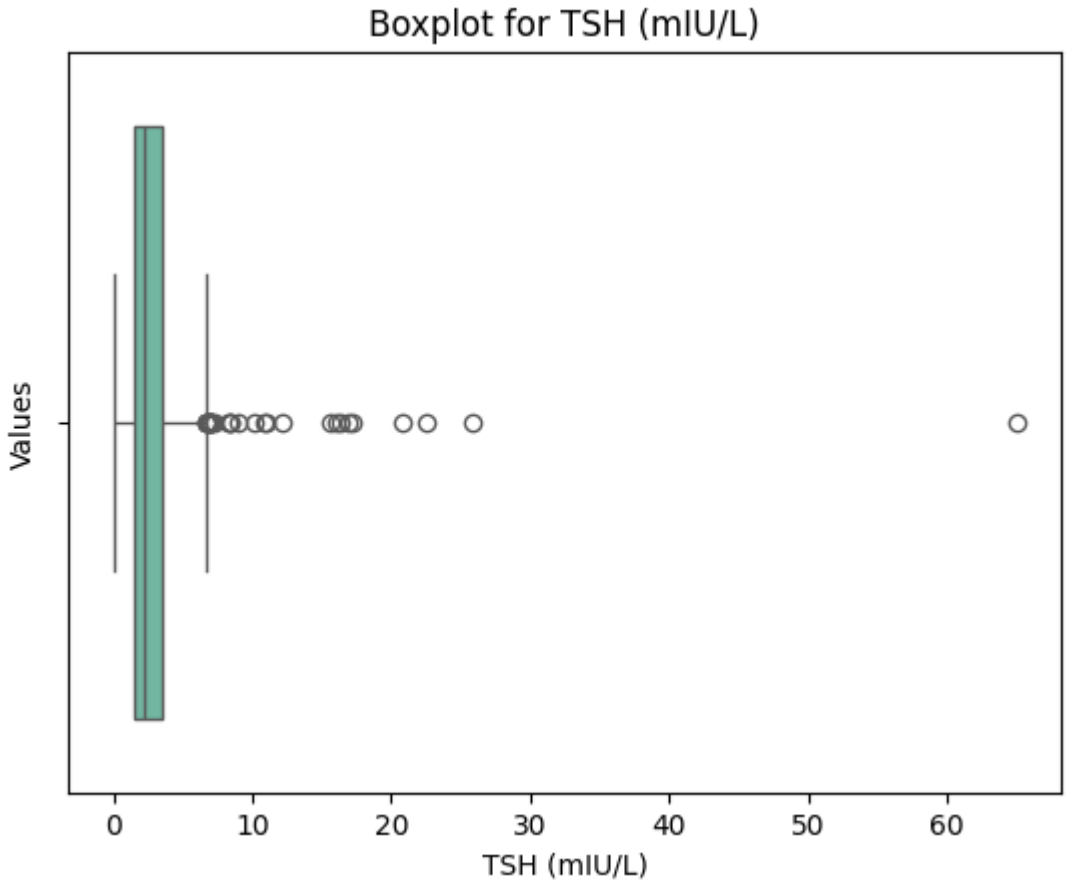
```
34. feature 16 (0.009147)   I   beta-HCG(mIU/mL)
35. feature 8 (0.004801) Pulse rate(bpm)
36. feature 32 (0.004153) Skin darkening (Y/N)
37. feature 36 (0.003891) Reg.Exercise(Y/N)
38. feature 35 (0.003615) Fast food (Y/N)
39. feature 13 (0.002861) Marraige Status (Yrs)
40. feature 14 (0.002200) Pregnant(Y/N)
41. feature 37 (0.001784) BP _Systolic (mmHg)
42. feature 19 (0.001082) LH(mIU/mL)
43. feature 22 (0.000895) Waist(inch)
44. feature 43 (0.000000) Endometrium (mm)
```

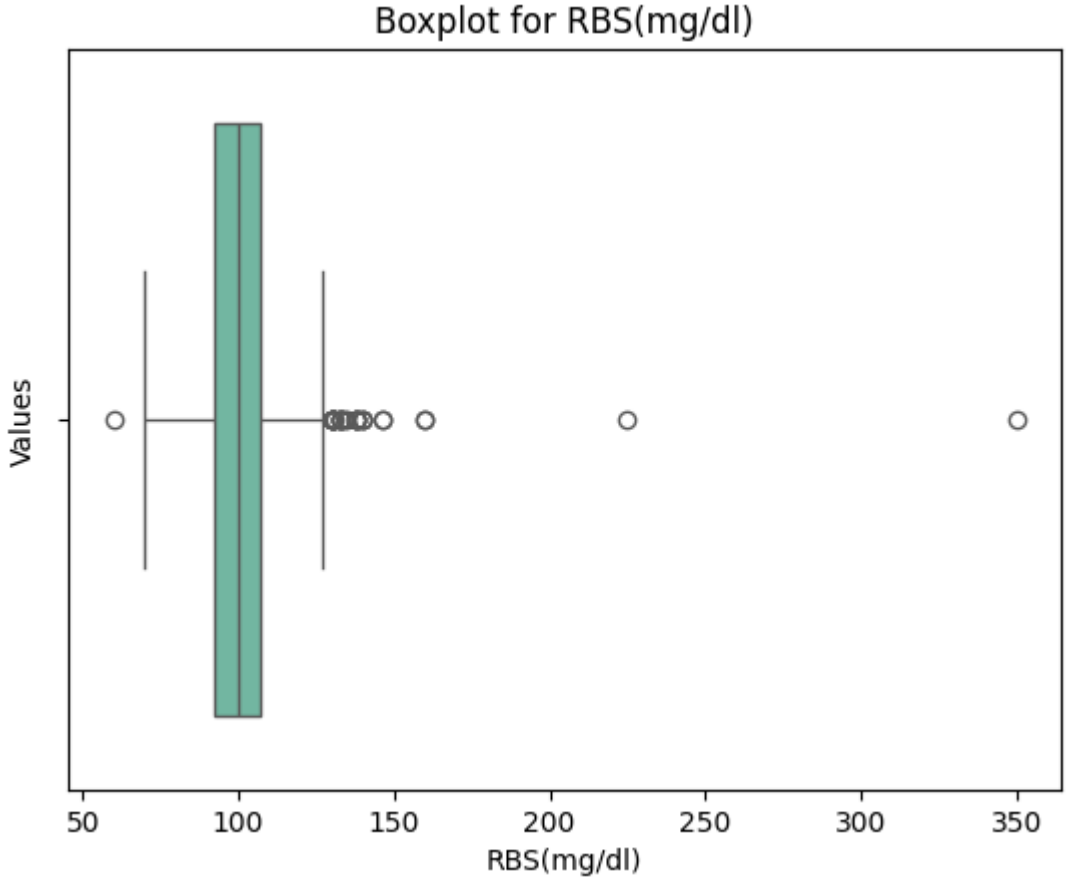


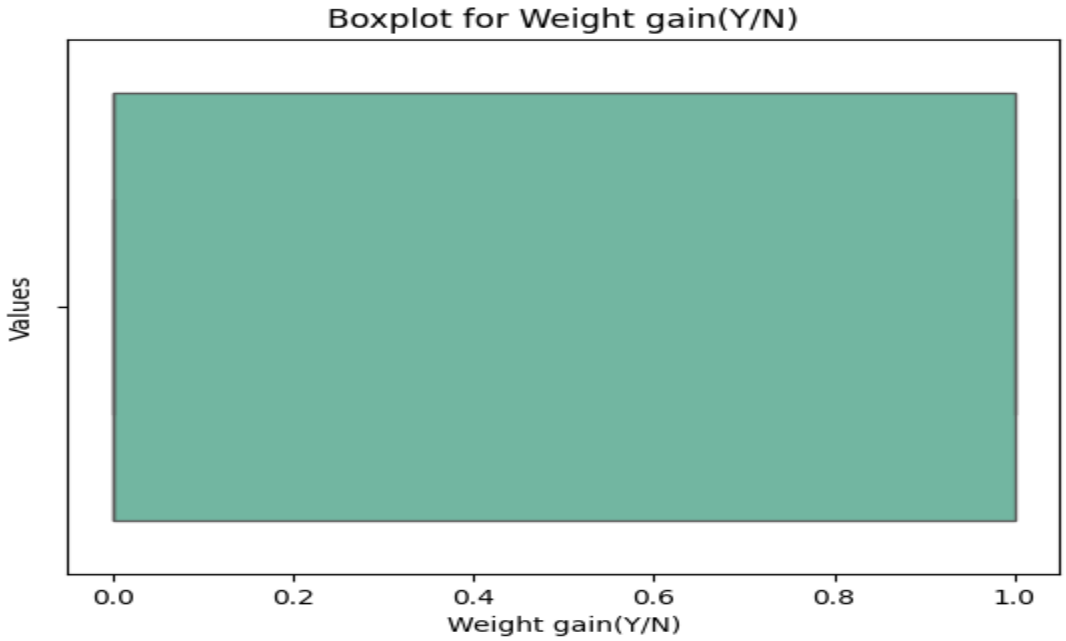

Boxplot for Hb(g/dl)
Values
Hb(g/dl)

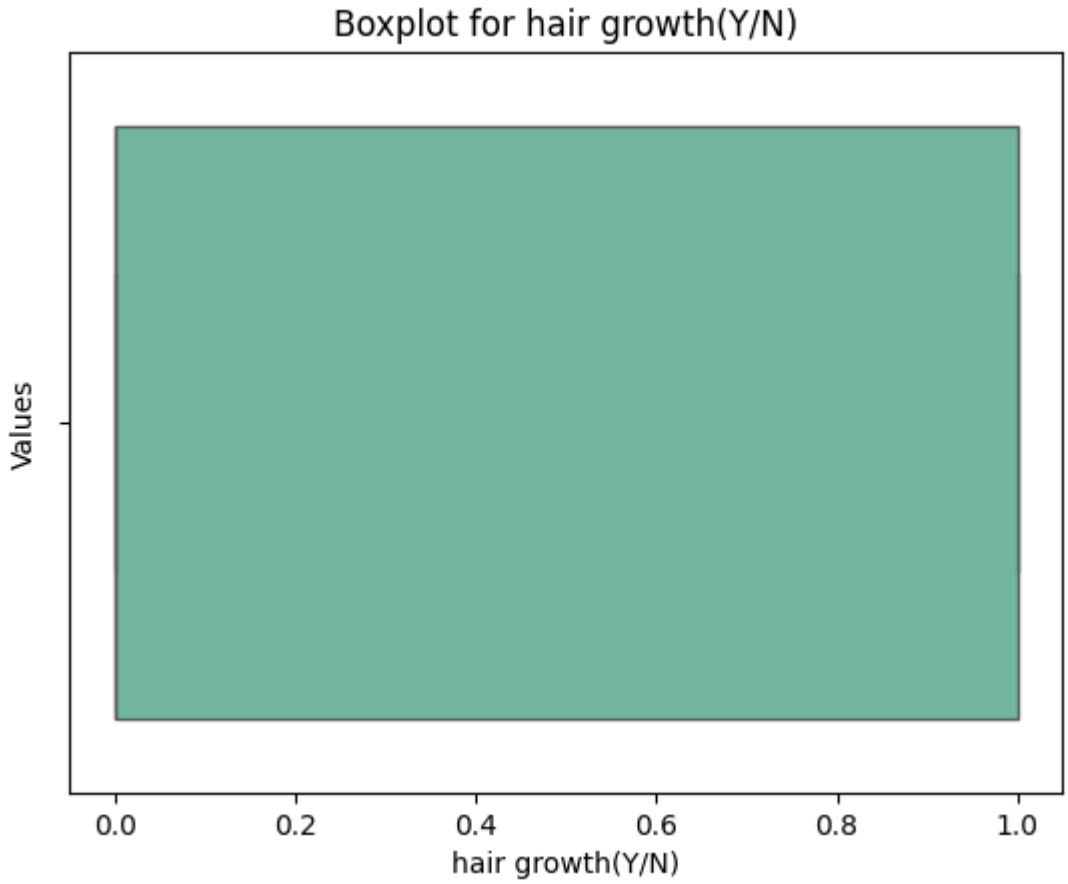

Boxplot for Cycle(R/I)
Values
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Cycle(R/I)

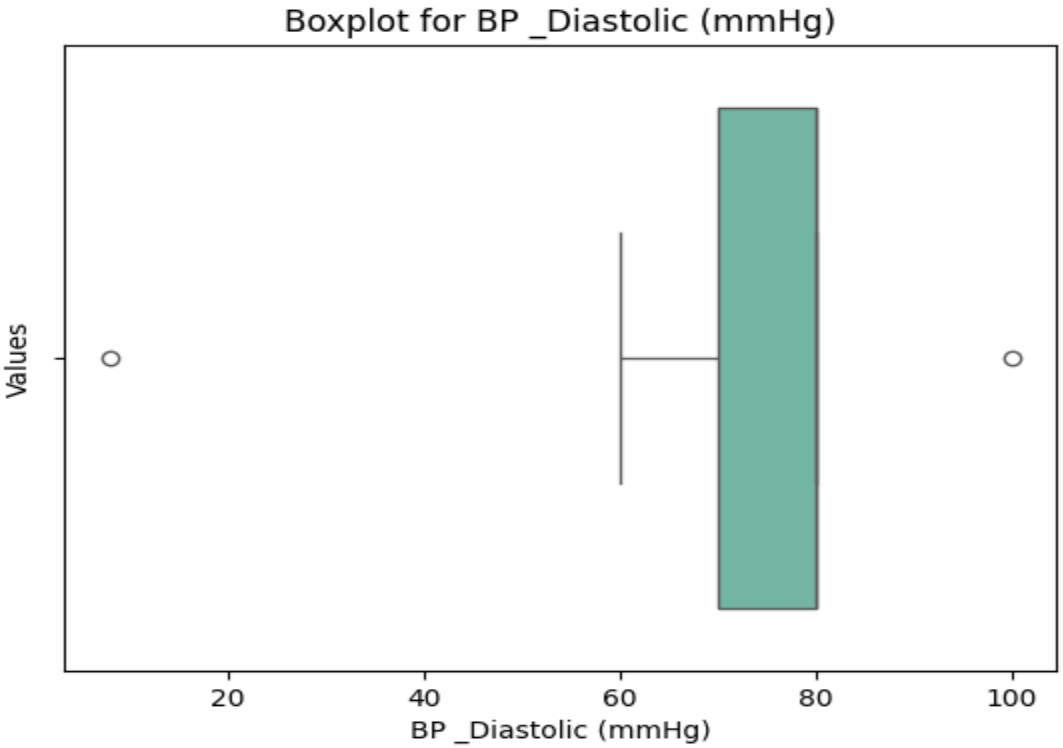

Boxplot for FSH(mIU/mL)
Values
0
1000
2000
3000
4000
5000
FSH(mIU/mL)

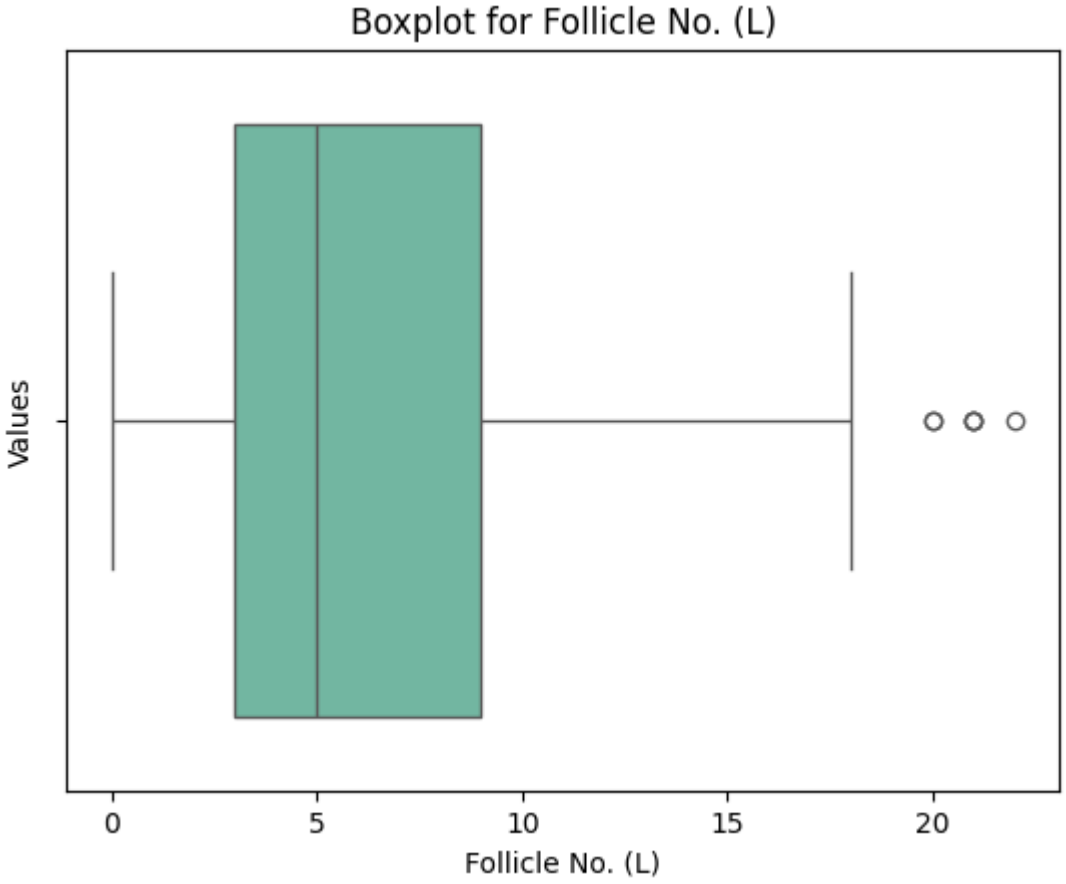

Boxplot for TSH (mIU/L)
Values
0
10
20
30
40
50
60
TSH (mIU/L)


Boxplot for RBS(mg/dl)
Values
50
100
150
200
250
300
350
RBS(mg/dl)

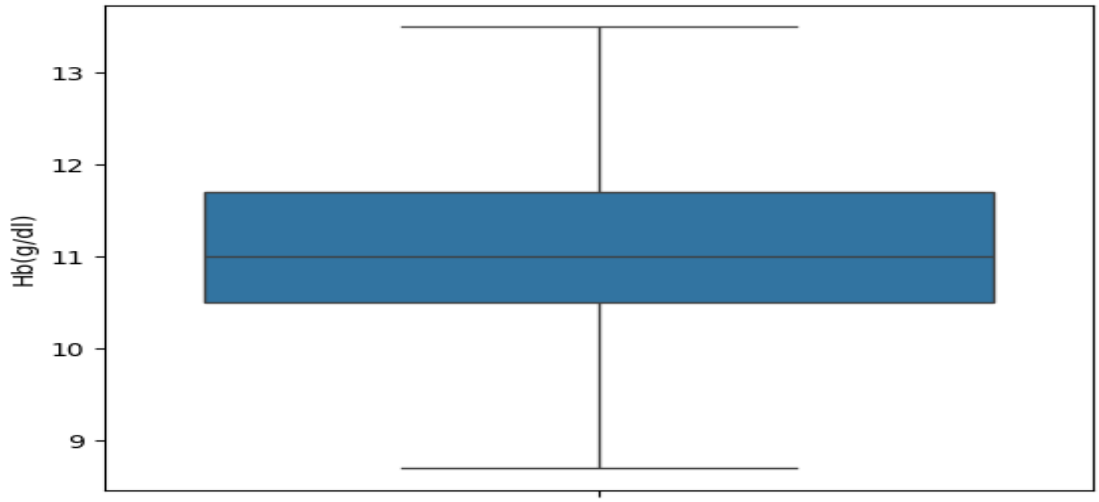

Boxplot for Weight gain(Y/N)
Values
0.0
0.2
0.4
0.6
0.8
1.0
Weight gain(Y/N)


Boxplot for hair growth(Y/N)
Values
0.0
0.2
0.4
0.6
0.8
1.0
hair growth(Y/N)

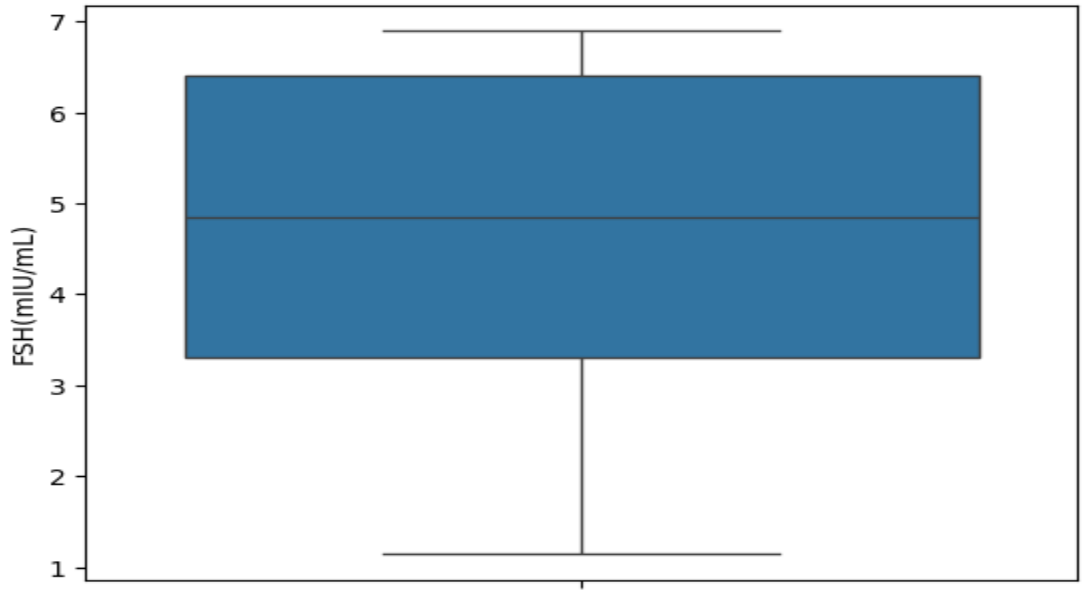

Boxplot for BP _Diastolic (mmHg)
Values
20
40
60
80
100
BP _Diastolic (mmHg)


Boxplot for Follicle No. (L)
Values
0
5
10
15
20
Follicle No. (L)

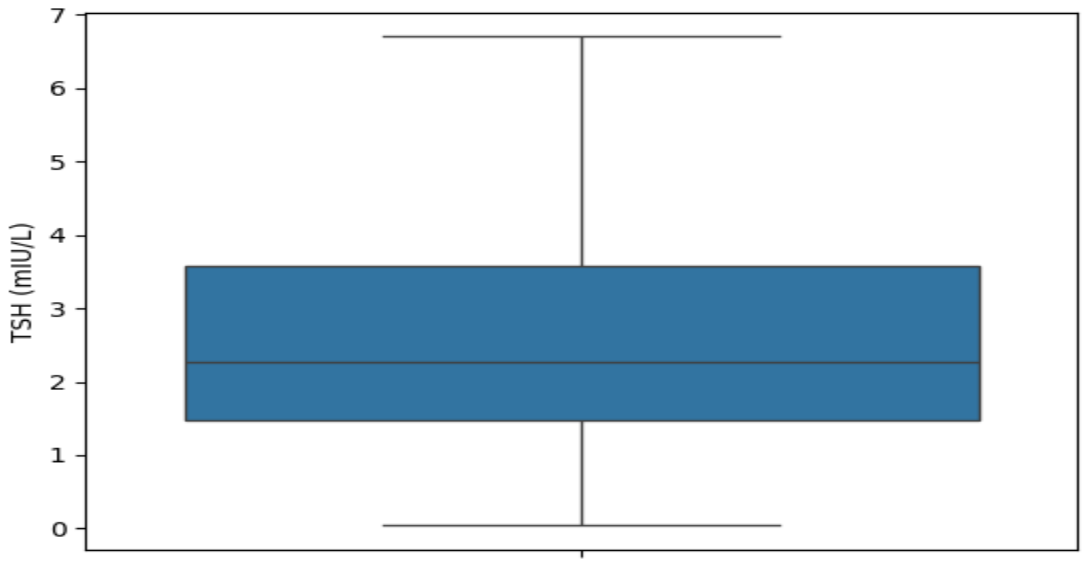
```
0.75    11.7
0.25    10.5
Name: Hb(g/dl), dtype: float64
11.7
10.5
1.1999999999999993
13.499999999999998
8.700000000000001
<Axes: ylabel='Hb(g/dl)'>
```



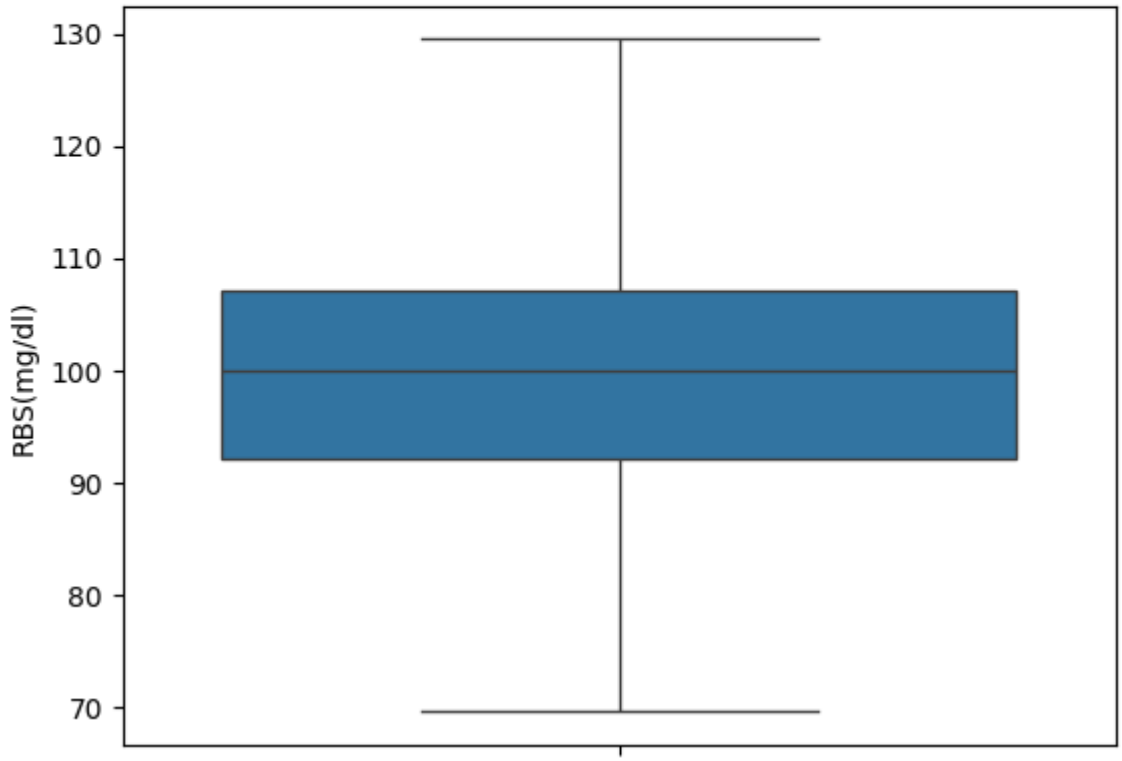

```
0.75    6.41
0.25    3.30
Name: FSH(mIU/mL), dtype: float64
6.41
3.3
1.1999999999999993
8.209999999999999
1.5000000000000009
<Axes: ylabel='FSH(mIU/mL)'>
```



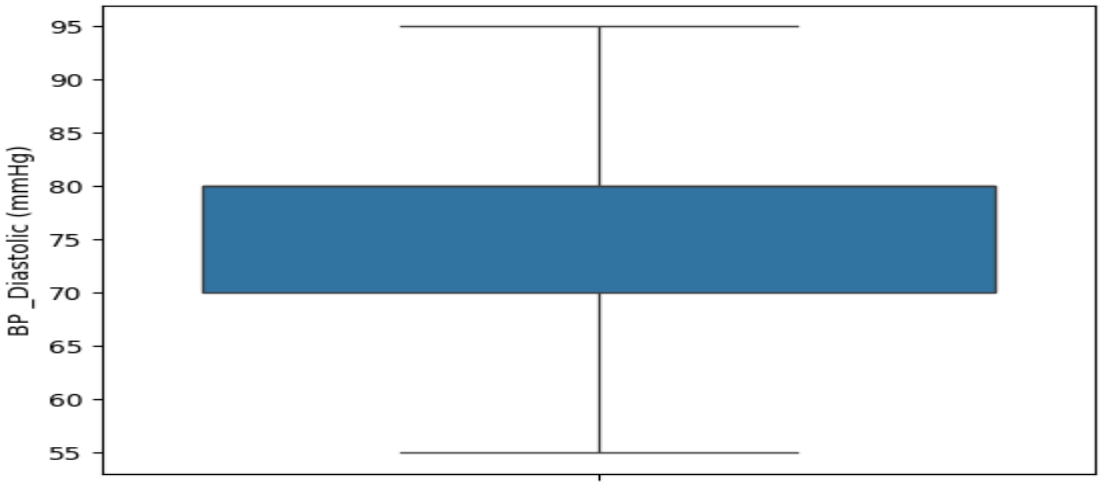
```
0.75    3.57
0.25    1.48
Name: TSH (mIU/L), dtype: float64
3.57
1.48
1.1999999999999993
5.369999999999999
-0.31999999999999895
<Axes: ylabel='TSH (mIU/L)'>
```



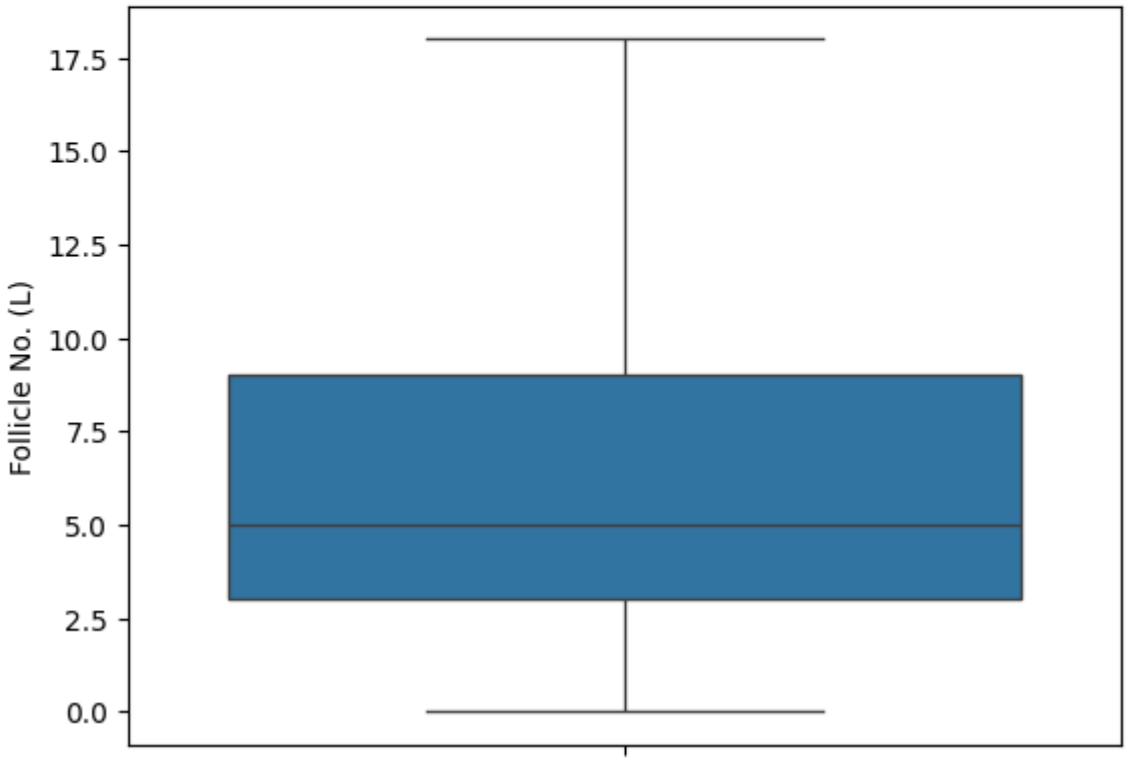

```
0.75    107.0
0.25     92.0
Name: RBS(mg/dl), dtype: float64
107.0
92.0
15.0
129.5
69.5
<Axes: ylabel='RBS(mg/dl)'>
```



```
0.75    80.0
0.25    70.0
Name: BP _Diastolic (mmHg), dtype: float64
80.0
70.0
15.0
102.5
47.5
<Axes: ylabel='BP _Diastolic (mmHg)'>
```



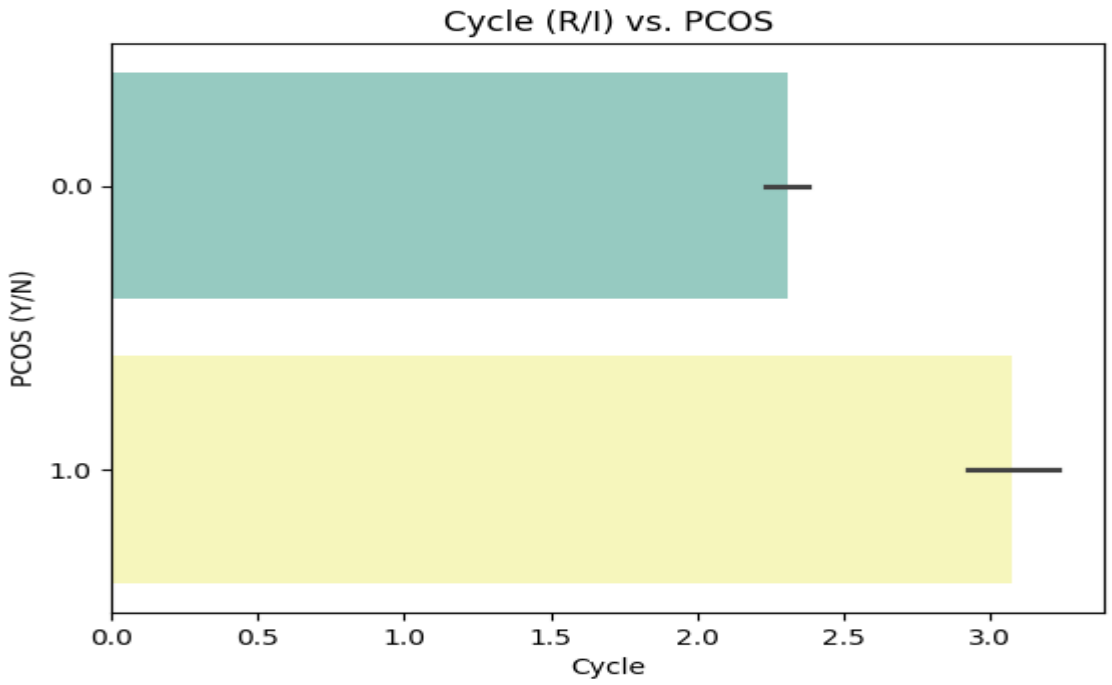

```
0.75    9.0
0.25    3.0
Name: Follicle No. (L), dtype: float64
9.0
3.0
6.0
18.0
-6.0
<Axes: ylabel='Follicle No. (L)'>
```

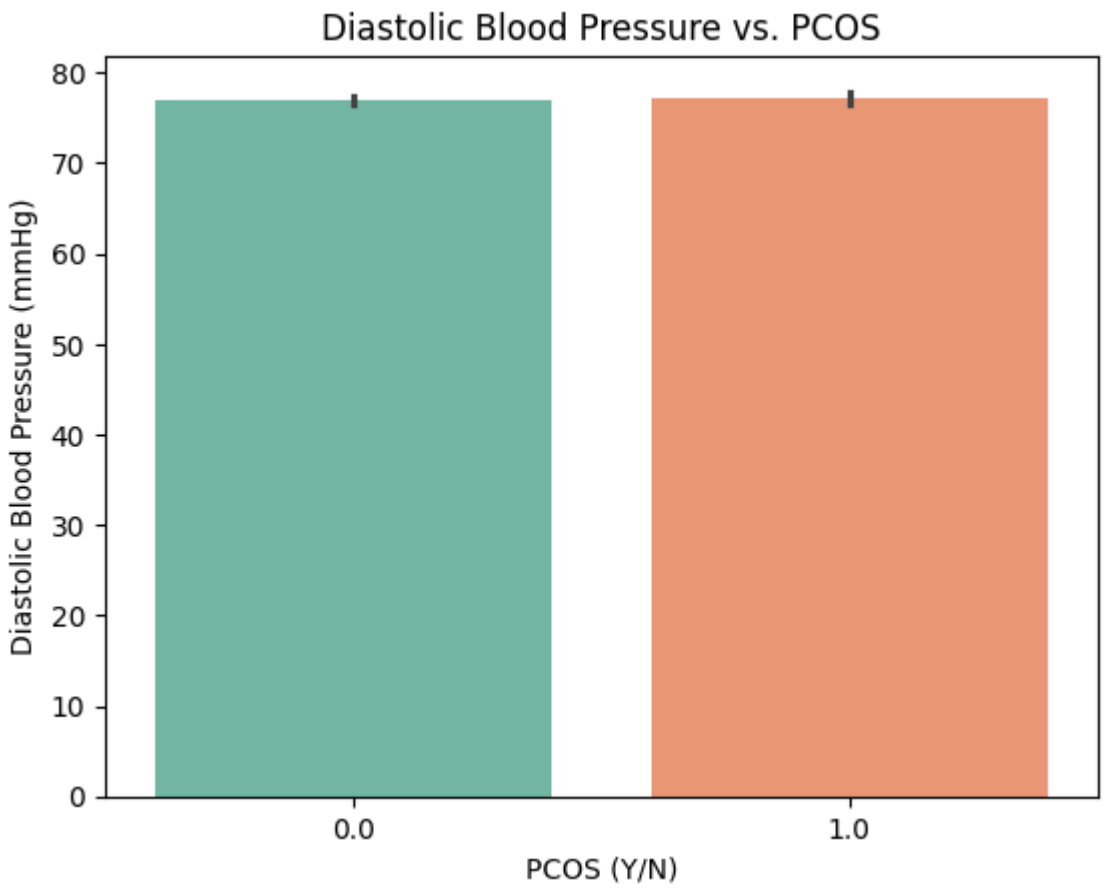


```
data.describe()
```

	Sl. No	Patient File No.	PCOS (Y/N)	Age (yrs)	Weight (Kg)	Height(Cm)	BMI	Blood Group	Pulse rate(bpm)	RR (breaths/min)	...	Fast food (Y/N)	Reg.Exercise(Y/N)
count	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	...	541.000000	541.000000
mean	271.000000	271.000000	0.327172	31.430684	59.637153	156.484835	23.929752	13.802218	73.247689	19.243993	...	0.514815	0.247689
std	156.317519	156.317519	0.469615	5.411006	11.028287	6.033545	2.447200	1.840812	4.430285	1.688629	...	0.499780	0.432070
min	1.000000	1.000000	0.000000	20.000000	31.000000	137.000000	15.100000	11.000000	13.000000	16.000000	...	0.000000	0.000000
25%	136.000000	136.000000	0.000000	28.000000	52.000000	152.000000	23.929752	13.000000	72.000000	18.000000	...	0.000000	0.000000
50%	271.000000	271.000000	0.000000	31.000000	59.000000	156.000000	23.929752	14.000000	72.000000	18.000000	...	1.000000	0.000000
75%	406.000000	406.000000	1.000000	35.000000	65.000000	160.000000	23.929752	15.000000	74.000000	20.000000	...	1.000000	0.000000
max	541.000000	541.000000	1.000000	48.000000	108.000000	180.000000	38.900000	18.000000	82.000000	28.000000	...	1.000000	1.000000

8 rows × 45 columns






```
y_train.value_counts()
```

```
PCOS (Y/N)
0.0    287
1.0    145
Name: count, dtype: int64
```

```
ybal.value_counts()
```

```
PCOS (Y/N)
1.0    287
0.0    287
Name: count, dtype: int64
```

```
model_= DecisionTreeClassifier()
model_.fit(X_train, y_train)
dt_pred = model_.predict(X_test)
acc_score = accuracy_score(y_test, dt_pred)
print("acc_score of model %.2f"%accuracy_score(y_test, dt_pred))
```

```
acc_score of model 0.80
```

```
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier(n_estimators=100)
my_model = rfc.fit(X_train, y_train)
#Making prediction and checking the test set
from sklearn.metrics import accuracy_score
pred_rfc = rfc.predict(X_test)
accuracy = accuracy_score(y_test, pred_rfc)
print(accuracy)
```

```
0.8715596330275229
```

```
from sklearn.metrics import accuracy_score
reg_model = LogisticRegression()
reg_model.fit(X_train, y_train)
reg_pred = reg_model.predict(X_test)
reg_acc_score = accuracy_score(y_test, reg_pred)
print("acc_score of model %.2f"%accuracy_score(y_test, reg_pred))
```

```
acc_score of model 0.88
```

x

	Hb(g/dl)	Cycle(R/I)	FSH(mIU/mL)	TSH (mIU/L)	RBS(mg/dl)	Weight gain(Y/N)	hair growth(Y/N)	BP _Diastolic (mmHg)	Follicle No. (L)
0	10.48	2.0	6.897407	0.680	92.0	0.0	0.0	80.0	3.0
1	11.70	2.0	6.730000	3.160	92.0	0.0	0.0	70.0	3.0
2	11.80	2.0	5.540000	2.540	84.0	0.0	0.0	80.0	13.0
3	12.00	2.0	6.897407	6.705	76.0	0.0	0.0	70.0	2.0
4	10.00	2.0	3.980000	3.570	84.0	0.0	0.0	80.0	3.0
...	...	...	...	...	...	...	...	...	...
536	11.00	2.0	6.897407	1.110	92.0	0.0	0.0	70.0	1.0
537	10.80	2.0	5.070000	2.050	108.0	1.0	0.0	70.0	9.0
538	10.80	2.0	6.897407	2.870	92.0	0.0	0.0	80.0	1.0
539	12.00	4.0	4.400000	2.500	115.0	0.0	0.0	70.0	7.0
540	10.20	4.0	3.990000	1.660	108.0	1.0	1.0	70.0	9.0

541 rows × 9 columns

```
y
```

```
0      0.0
1      0.0
2      1.0
3      0.0
4      0.0
      ...
536    0.0
537    0.0
538    0.0
539    0.0
540    1.0
Name: PCOS (Y/N), Length: 541, dtype: float64
```

```
print(model_.predict([[10.20,4.0,3.990000,1.660,108.0,1.0,1.0,70.0,9.0]]))
```

```
[1.]
```

```
from sklearn.metrics import classification_report
print(classification_report (reg_pred, y_test))
```

```
              precision    recall  f1-score   support

         0.0       0.96      0.88      0.92        84
         1.0       0.69      0.88      0.77        25

    accuracy                           0.88       109
   macro avg       0.82      0.88      0.85       109
weighted avg       0.90      0.88      0.89       109
```

```
print(classification_report(dt_pred, y_test))
```

```
              precision    recall  f1-score   support

         0.0       0.87      0.85      0.86        79
         1.0       0.62      0.67      0.65        30

    accuracy                           0.80       109
   macro avg       0.75      0.76      0.75       109
weighted avg       0.80      0.80      0.80       109
```

```
from sklearn.metrics import classification_report
classi_report = classification_report (y_test, pred_rfc)
print(classi_report)
```

```
              precision    recall  f1-score   support

         0.0       0.87      0.96      0.91        77
         1.0       0.88      0.66      0.75        32

    accuracy                           0.87       109
   macro avg       0.87      0.81      0.83       109
weighted avg       0.87      0.87      0.87       109
```

```
print(classification_report (reg_pred, y_test))
```

```
              precision    recall  f1-score   support

         0.0       0.96      0.88      0.92        84
         1.0       0.69      0.88      0.77        25

    accuracy                           0.88       109
   macro avg       0.82      0.88      0.85       109
weighted avg       0.90      0.88      0.89       109
```