Web Project (MDS481) SRS

LunaCare: Comprehensive Website for Women's Menstrual Health and Hygiene

Submitted by

Asmita Mondal (2348018)

Swarnasish Banerjee (2348066)

Sayan Pal (2349056)

Department of Statistics and Data Science

July 2024

Objective and Problem Description

Objective

The primary goal of LunaCare is to create a comprehensive, user-friendly website that serves as a trusted resource for women's menstrual health and hygiene. The platform will provide educational content, interactive tools, and a community forum to support women in managing their menstrual health effectively.

Problem Description

Why: Women's menstrual health is often overlooked, leading to a lack of accessible, reliable, and comprehensive resources. Many women struggle to find trustworthy information and tools to manage their menstrual health, which can impact their overall well-being.

What: LunaCare aims to bridge this gap by offering a dedicated platform that focuses exclusively on menstrual health. The website will provide educational content, practical tools, and a supportive community to help women navigate their menstrual health journey.

Beneficiaries: Women of all ages who seek reliable information and support for managing their menstrual health will benefit from LunaCare. This includes teenagers, young adults, women with conditions like PCOS (Polycystic Ovary Syndrome), and those approaching menopause. The data used for training all algorithms are sourced from reliable medical databases hence making it a strong source of support for any time help.

Project Study

Existing System and their Limitations

Currently, there are a few websites and mobile apps that offer period calculators and general health information for women. However, these platforms have significant limitations:

- **Lack of Specificity**: There is no dedicated website that exclusively addresses menstrual health. The available resources are often part of broader women's health websites, which can dilute the focus on menstrual health.
- **Insufficient Support for PCOS**: Conditions like PCOS, which affect menstrual health, are not adequately covered in existing platforms. Women with PCOS often find it challenging to get tailored information and support.**No Real-Time Data Insights**: There is a lack of data-backed insights and analysis on menstrual health trends and issues.
- **Absence of Community Support**: No platform provides a space for women to connect, share experiences, and support each other.

Proposed System

LunaCare aims to address the limitations of existing systems by providing a comprehensive platform dedicated to menstrual health. The proposed system will include:

- **Educational Content**: In-depth articles, videos, and infographics covering various aspects of menstrual health, including menstrual cycle phases, common issues, and tips for managing menstrual health.
- **Interactive Tools**: Period trackers, and fertility calculators to help women monitor and manage their menstrual health.
- **Community Forum**: A supportive community space where women can share experiences, ask questions, and offer support to each other.
- **PCOS Support**: Dedicated resources and tools for women with PCOS, including information on managing symptoms and connecting with others facing similar challenges.
- **Real-Time Data Insights**: Awareness blogs and reports backed by exploratory data analysis (EDA) of real-time data to provide a better understanding of menstrual health trends and issues worldwide.
- **Chat Assistance:** A chatbot solely dedicated to answering menstruation related questions for curious minds.

Benefits of Proposed System

- **Focused Approach**: By dedicating the platform exclusively to menstrual health, LunaCare ensures that users receive targeted and relevant information.
- **Comprehensive Resources**: A wide range of educational content and tools helps women manage their menstrual health effectively.
- **Community Support**: The community forum fosters a sense of belonging and support, helping women connect with others who share similar experiences.
- **Enhanced PCOS Support**: Tailored resources and tools for women with PCOS address a significant gap in existing systems.
- **Data-Driven Insights**: Real-time data analysis and awareness blogs provide valuable insights into menstrual health trends and issues, helping users make informed decisions.

Literature Review

Year of Implementation	Author/Company	Techniques/ Algorithms	Gaps/Drawbacks
2002	Andrew S. Rowland, Donna Day Baird, Stuart Long, Ganesa Wegienka, Siobán D. Harlow, Michael Alavanja, Dale P. Sandler (Published in Epidemiology)	Cross-sectional data analysis, Logistic regression	The study's findings were crucial in understanding medical and lifestyle factors affecting the menstrual cycle, which informed the creation of the chatbot application. However, the reliance on data from specific regions (Iowa and North Carolina) and the agricultural health context may limit generalizability.
2018	Jinju Bae, Susan Park, and Jin-Won Kwon	Univariate and multiple logistic regression analyses, logit and linear models	This paper was used as a reference to create a chatbot application and understand the validity of the project. It was necessary to research the factors in order to create a proper website that is reliable and informative.
2019	Amsy Denny, Anita Raj, Ashi Ashok, C Maneesh Ram, Remya George (Published by IEEE)	Naïve Bayes, Logistic Regression, K-Nearest Neighbor (KNN), Classification and Regression Trees (CART), Random Forest Classifier (RFC), Support Vector Machine (SVM)	The accuracy of the Random Forest Classifier, while the highest among the tested algorithms, is 89.02%, which may be insufficient for medical domains where accuracy greater than 95% is often necessary. Additionally, the study does not account for

			potential biases in the dataset, and the reliance on PCA for feature transformation might overlook other important features.
2021	Shakoor Ahmad Bhat (School of Computing, National College of Ireland)	XGBRF, CatBoost, SMOTE, Gradient Boosting, Random Forest, Logistic Regression, HRFLR, SVM, Decision Tree, MLP	The reliance on synthetic data resampling (SMOTE) might introduce biases, and the study does not address the potential impact of these biases on the model's real-world applicability. Additionally, the comparison with baseline approaches does not consider the computational complexity and scalability of the proposed models.
2022	Wenwen Xu, Zheng Zhu, Quan Wang, Jing Jin, Haiyang Zhao, Fang Shao, Qingling Ren, Hui Wang (Jiangsu Provincial Hospital of Traditional Chinese Medicine, Nanjing Medical University)	OvAge calculator, WeChat mini program, Generalized Linear Model (GLM)	The reliance on self-reporting through a WeChat mini program may introduce biases due to inaccurate or incomplete self-reported data. Additionally, the model's applicability is limited to Chinese women, and it may not be generalizable to other populations.

Requirements Specification

Functional Requirements

PCOS Predictor

1. **PCOS Predictor for Medical Professionals**:
    - **Input**: Medical parameters from reports, such as hormone levels, ultrasound results, and other relevant data.
    - **Process**: The system will analyze the input data using an algorithm designed to predict the likelihood of PCOS.
    - **Output**: A report indicating the probability of the patient having PCOS.
2. **PCOS Quiz for General Users**:
    - **Input**: User responses to a series of questions about symptoms and menstrual health.
    - **Process**: The system will evaluate the answers using a predefined scoring system to determine the likelihood of PCOS.
    - **Output**: A result indicating whether the user shows signs of PCOS and suggesting whether they should consult a medical professional.

Period Calculator

- **Input**: Last period date and average cycle length.
- **Process**: The system will use the input data to calculate the predicted date of the user's next period.
- **Output**: The next expected period date.

Chatbot

- **Input**: User queries related to menstruation and menstrual health myths.
- **Process**: The chatbot, trained on over 450 questions, will process the input using natural language processing (NLP) to understand and respond appropriately.
- **Output**: Relevant and informative responses to user queries.

Blog Page

- **Input**: Topics for blog posts and relevant datasets.
- **Process**: The system will create blog posts that spread awareness about menstrual health issues. Each blog will be backed by analysis of real-time data to provide accurate and insightful information.
- **Output**: Informative blog posts with data-backed insights on various menstrual health topics.

Non-Functional Requirements

System Constraints

- **Limited CPU/GPU Access**: The system will operate under constraints of limited CPU and GPU resources, ensuring efficient usage of available computational power.
- **Scalability**: The platform should be able to handle an increasing number of users without significant degradation in performance.

Response Time

- **PCOS Predictor**: The prediction results should be generated within a reasonable timeframe, ideally under a minute for medical professionals and instantaneously for quiz results.
- **Period Calculator**: The next period date should be calculated and displayed within seconds.
- **Chatbot**: Responses should be generated in real-time, providing instant feedback to user queries.

Security

- **Data Encryption**: All user data, including medical parameters and quiz responses, will be encrypted to ensure privacy and security.
- **Chatbot Conversations**: Conversations with the chatbot will be encrypted and stored in secure databases to protect user privacy.

Reusability

- **Modular Design**: The system will be designed with modularity in mind, allowing components (e.g., PCOS predictor, period calculator, chatbot) to be reused or updated independently.
- **API Integration**: The platform will support integration with other systems and tools via APIs, enhancing its functionality and interoperability.

Modifiability

- **Scalable Architecture**: The system architecture will be designed to allow easy modifications and updates, accommodating new features and improvements as needed.
- **User Feedback**: The platform will incorporate user feedback mechanisms to continuously improve and adapt to users' needs.

Software and Hardware Requirements

Software

- **Python**: Programming language with necessary libraries installed (e.g., Pandas, NumPy, Scikit-learn, NLTK).
- **Kaggle Notebook**: For development and testing environments.
- **TensorFlow**: GPU accelerated for machine learning tasks.

Hardware

- **RAM**: 8 GB or above.
- **CPU**: 8th Generation Intel® Core™ i3 Processor or above.
- **GPU**: 8 GB or above.
- **OS**: 64-bit operating system.
- **Graphics Card**: Integrated.

Dataset Description For Blogs

Menstrual Cup Dataset Description

This dataset includes detailed information about various menstrual cups, which are reusable feminine hygiene products. Each entry in the dataset represents a different menstrual cup model and contains the following attributes:

1. **Brand**:
    - **Description**: The name of the manufacturer or brand of the menstrual cup.
    - **Example**: "DivaCup", "Lunette", "Mooncup"
2. **Size**:
    - **Description**: The size of the menstrual cup, typically denoted as small, medium, large, or using numerical values.
    - **Example**: "Small", "Large", "1", "2"
3. **Length**:
    - **Description**: The total length of the menstrual cup in millimeters (mm).
    - **Example**: "70 mm"
4. **Diameter**:
    - **Description**: The diameter of the menstrual cup's rim in millimeters (mm).
    - **Example**: "45 mm"
5. **Capacity (To Holes)**:
    - **Description**: The volume capacity of the menstrual cup up to the air holes, measured in milliliters (ml).
    - **Example**: "20 ml"
6. **Capacity (Listed)**:
    - **Description**: The total volume capacity of the menstrual cup as listed by the manufacturer, measured in milliliters (ml).
    - **Example**: "25 ml"
7. **Stem**:
    - **Description**: The type and length of the stem of the menstrual cup, which is used for removal.
    - **Example**: "Flat", "Ball", "Ring", "15 mm"
8. **Total Length**:
    - **Description**: The combined length of the menstrual cup and its stem in millimeters (mm).
    - **Example**: "85 mm"
9. **Firmness**:
    - **Description**: The firmness of the menstrual cup, usually described as soft, medium, or firm.
    - **Example**: "Soft", "Firm"
10. **Coupon Available**:
    - **Description**: Indicates whether there are any promotional coupons available for the menstrual cup.

Menstrual Cycle Dataset Description

This dataset contains detailed information about menstrual cycles and various related attributes for patients, including demographic and reproductive details. Each entry in the dataset represents a specific menstrual cycle for an individual and includes the following attributes:

1. **CycleNumber:**
    - **Description**: Unique identifier for each menstrual cycle record.
    - **Example**: "C12345"
2. **Group:**
    - **Description**: The group or category to which the patient belongs.
    - **Example**: "Control", "Treatment"
3. **CycleWithPeakOrNot:**
    - **Description**: Indicates if the cycle had a peak day (Yes/No).
    - **Example**: "Yes"
4. **ReproductiveCategory:**
    - **Description**: The reproductive status of the patient.
    - **Example**: "Fertile", "Infertile", "Pregnant"
5. **LengthofCycle:**
    - **Description**: Length of the menstrual cycle in days.
    - **Example**: "28"
6. **MeanCycleLength:**
    - **Description**: Average length of menstrual cycles.
    - **Example**: "27.5"
7. **EstimatedDayofOvulation:**
    - **Description**: Estimated day of ovulation in the cycle.
    - **Example**: "14"
8. **LengthofLutealPhase:**
    - **Description**: Length of the luteal phase in days.
    - **Example**: "14"
9. **FirstDayofHigh:**
    - **Description**: First day of high fertility.
    - **Example**: "10"
10. **TotalNumberofHighDays:**
    - **Description**: Total number of high fertility days in the cycle.
    - **Example**: "5"
11. **TotalHighPostPeak:**
    - **Description**: Total number of high fertility days after the peak day.
    - **Example**: "2"
12. **TotalNumberofPeakDays:**
    - **Description**: Total number of peak fertility days.
    - **Example**: "3"
13. **TotalDaysofFertility:**
    - **Description**: Total number of days with fertility indicators.

14. **TotalFertilityFormula:**
    - **Description**: Calculated formula indicating total fertility.
    - **Example**: "TF8"
15. **LengthofMenses:**
    - **Description**: Length of the menstrual bleeding phase in days.
    - **Example**: "5"
16. **MeanMensesLength:**
    - **Description**: Average length of menstrual bleeding across cycles.
    - **Example**: "4.5"
17. **MensesScoreDayOne** to **MensesScoreDayFifteen:**
    - **Description**: Intensity score of menstrual bleeding for each day (1-15).
    - **Example**: "3" for Day One, "2" for Day Two, etc.
18. **TotalMensesScore:**
    - **Description**: Sum of menses scores over the cycle.
    - **Example**: "10"
19. **MeanBleedingIntensity:**
    - **Description**: Average bleeding intensity score.
    - **Example**: "2"
20. **NumberofDaysofIntercourse:**
    - **Description**: Total number of days with intercourse.
    - **Example**: "4"
21. **IntercourseInFertileWindow:**
    - **Description**: Indicates if intercourse occurred during the fertile window (Yes/No).
    - **Example**: "Yes"
22. **UnusualBleeding:**
    - **Description**: Indicates if there was unusual bleeding during the cycle (Yes/No).
    - **Example**: "No"
23. **PhasesBleeding:**
    - **Description**: Phase(s) during which bleeding occurred (e.g., "Luteal", "Follicular").
    - **Example**: "Follicular"
24. **IntercourseDuringUnusBleed:**
    - **Description**: Indicates if intercourse occurred during unusual bleeding (Yes/No).
    - **Example**: "No"
25. **Age:**
    - **Description**: Age of the patient.
    - **Example**: "30"
26. **Maristatus:**
    - **Description**: Marital status of the patient.
    - **Example**: "Married"
27. **MaristatusM:**
    - **Description**: Marital status of the patient's partner.
    - **Example**: "Married"

28. **Yearsmarried:**
    - **Description**: Number of years married.
    - **Example**: "5"
29. **Wedding:**
    - **Description**: Wedding date.
    - **Example**: "2015-06-20"
30. **Religion:**
    - **Description**: Religion of the patient.
    - **Example**: "Christian"
31. **ReligionM:**
    - **Description**: Religion of the patient's partner.
    - **Example**: "Christian"
32. **Ethnicity:**
    - **Description**: Ethnicity of the patient.
    - **Example**: "Asian"
33. **EthnicityM:**
    - **Description**: Ethnicity of the patient's partner.
    - **Example**: "Caucasian"
34. **Schoolyears:**
    - **Description**: Number of years of education.
    - **Example**: "16"
35. **SchoolyearsM:**
    - **Description**: Number of years of education for the partner.
    - **Example**: "18"
36. **OccupationM:**
    - **Description**: Occupation of the patient's partner.
    - **Example**: "Engineer"
37. **IncomeM:**
    - **Description**: Income of the patient's partner.
    - **Example**: "$80,000"
38. **Height:**
    - **Description**: Height of the patient in centimeters.
    - **Example**: "165"
39. **Weight:**
    - **Description**: Weight of the patient in kilograms.
    - **Example**: "60"

Dataset Description For PCOS Detection

PCOS (Polycystic Ovary Syndrome) Dataset Description

This dataset contains detailed medical and physiological information for patients being evaluated for Polycystic Ovary Syndrome (PCOS). Each entry in the dataset represents an individual patient's record and includes the following attributes:

1. **Patient File No.:**
    - **Description**: Unique identifier for each patient's file.
    - **Example**: "PF12345"
2. **PCOS (Y/N):**
    - **Description**: Indicates whether the patient has been diagnosed with PCOS (Yes/No).
    - **Example**: "Y"
3. **Age (yrs):**
    - **Description**: Age of the patient in years.
    - **Example**: "28"
4. **Weight (Kg):**
    - **Description**: Weight of the patient in kilograms.
    - **Example**: "68"
5. **Height (Cm):**
    - **Description**: Height of the patient in centimeters.
    - **Example**: "165"
6. **BMI:**
    - **Description**: Body Mass Index of the patient.
    - **Example**: "24.9"
7. **Blood Group:**
    - **Description**: Blood group of the patient.
    - **Example**: "O+"
8. **Pulse rate (bpm):**
    - **Description**: Pulse rate of the patient in beats per minute.
    - **Example**: "72"
9. **RR (breaths/min):**
    - **Description**: Respiratory rate of the patient in breaths per minute.
    - **Example**: "16"
10. **Hb (g/dl):**
    - **Description**: Hemoglobin level in grams per deciliter.
    - **Example**: "13.5"
11. **Cycle (R/I):**
    - **Description**: Menstrual cycle regularity (Regular/Irregular).
    - **Example**: "R"
12. **Cycle length (days):**
    - **Description**: Length of the menstrual cycle in days.
    - **Example**: "28"
13. **Marriage Status (Yrs):**
    - **Description**: Number of years married.
    - **Example**: "5"
14. **Pregnant (Y/N):**
    - **Description**: Indicates if the patient is currently pregnant (Yes/No).
    - **Example**: "N"

15. **No. of abortions:**
    - **Description**: Number of abortions the patient has had.
    - **Example**: "1"
16. **I beta-HCG (mIU/mL):**
    - **Description**: Level of beta-HCG hormone in the first test, measured in milli-international units per milliliter.
    - **Example**: "0.5"
17. **II beta-HCG (mIU/mL):**
    - **Description**: Level of beta-HCG hormone in the second test, measured in milli-international units per milliliter.
    - **Example**: "0.6"
18. **FSH (mIU/mL):**
    - **Description**: Follicle-stimulating hormone level in milli-international units per milliliter.
    - **Example**: "6.8"
19. **LH (mIU/mL):**
    - **Description**: Luteinizing hormone level in milli-international units per milliliter.
    - **Example**: "8.5"
20. **FSH/LH:**
    - **Description**: Ratio of FSH to LH.
    - **Example**: "0.8"
21. **Hip (inch):**
    - **Description**: Hip circumference in inches.
    - **Example**: "38"
22. **Waist (inch):**
    - **Description**: Waist circumference in inches.
    - **Example**: "32"
23. **Waist Ratio:**
    - **Description**: Ratio of waist circumference to hip circumference.
    - **Example**: "0.84"
24. **TSH (mIU/L):**
    - **Description**: Thyroid-stimulating hormone level in milli-international units per liter.
    - **Example**: "2.5"
25. **AMH (ng/mL):**
    - **Description**: Anti-Müllerian hormone level in nanograms per milliliter.
    - **Example**: "3.2"
26. **PRL (ng/mL):**
    - **Description**: Prolactin level in nanograms per milliliter.
    - **Example**: "12"
27. **Vit D3 (ng/mL):**
    - **Description**: Vitamin D3 level in nanograms per milliliter.
    - **Example**: "30"

28. **PRG (ng/mL):**
    - **Description**: Progesterone level in nanograms per milliliter.
    - **Example**: "1.5"
29. **RBS (mg/dl):**
    - **Description**: Random blood sugar level in milligrams per deciliter.
    - **Example**: "110"
30. **Weight gain (Y/N):**
    - **Description**: Indicates if the patient has experienced weight gain (Yes/No).
    - **Example**: "Y"
31. **Hair growth (Y/N):**
    - **Description**: Indicates if the patient has experienced excessive hair growth (Yes/No).
    - **Example**: "Y"
32. **Skin darkening (Y/N):**
    - **Description**: Indicates if the patient has experienced skin darkening (Yes/No).
    - **Example**: "N"
33. **Hair loss (Y/N):**
    - **Description**: Indicates if the patient has experienced hair loss (Yes/No).
    - **Example**: "Y"
34. **Pimples (Y/N):**
    - **Description**: Indicates if the patient has experienced pimples (Yes/No).
    - **Example**: "Y"
35. **Fast food (Y/N):**
    - **Description**: Indicates if the patient frequently consumes fast food (Yes/No).
    - **Example**: "N"
36. **Reg. Exercise (Y/N):**
    - **Description**: Indicates if the patient engages in regular exercise (Yes/No).
    - **Example**: "Y"
37. **BP Systolic (mmHg):**
    - **Description**: Systolic blood pressure in millimeters of mercury.
    - **Example**: "120"
38. **BP Diastolic (mmHg):**
    - **Description**: Diastolic blood pressure in millimeters of mercury.
    - **Example**: "80"
39. **Follicle No. (L):**
    - **Description**: Number of follicles in the left ovary.
    - **Example**: "10"
40. **Follicle No. (R):**
    - **Description**: Number of follicles in the right ovary.
    - **Example**: "12"
41. **Avg. F size (L) (mm):**
    - **Description**: Average size of follicles in the left ovary in millimeters.
    - **Example**: "8"

42. **Avg. F size (R) (mm):**
    - **Description**: Average size of follicles in the right ovary in millimeters.
    - **Example**: "9"
43. **Endometrium (mm):**
    - **Description**: Thickness of the endometrium in millimeters.
    - **Example**: "7"

Comprehensive Women's Health Survey

Description

This survey is designed to gather detailed information about various health aspects of women, particularly focusing on menstruation, physical characteristics, lifestyle habits, and health conditions. The collected data will help in understanding patterns and correlations related to women's health issues, especially those linked to Polycystic Ovary Syndrome (PCOS) and Polycystic Ovary Disorder (PCOD).

Questions Overview

1. **Age (in Years)**

   - **Label:** Age

   - **Description:** Enter your age in completed years.

2. **Weight (in Kg)**

   - **Label:** Weight

   - **Description:** Enter your weight in kilograms.

3. **Height (in Cm / Feet)**

   - **Label**: Height

   - **Description:** Enter your height in centimeters or feet.

4. ***Can you tell us your blood group?***

- ***Label***: BloodGroup

- ***Description*:** Indicate your blood group (e.g., A+, B-, O+).

5. ***After how many months do you get your periods? (select 1 if every month/regular)***

- ***Label*:** PeriodFrequency

- ***Description*:** Enter the number of months between your periods. Select 1 if they occur every month (regular).

6. ***Have you gained weight recently?***

- ***Label*:** WeightGain

- ***Description*:** Indicate if you have experienced recent weight gain (Yes/No).

7. ***Do you have excessive body/facial hair growth?***

- ***Label*:** HairGrowth

- ***Description*:** Indicate if you have noticed excessive hair growth on your body or face (Yes/No).

8. ***Are you noticing skin darkening recently?***

- ***Label*:** SkinDarkening

- ***Description*:** Indicate if you have noticed any recent skin darkening (Yes/No).

9. ***Do you have hair loss/hair thinning/baldness?***

- ***Label*:** HairLoss

- ***Description*:** Indicate if you are experiencing hair loss, thinning, or baldness (Yes/No).

10. ***Do you have pimples/acne on your face/jawline?***

- ***Label*:** Acne

- ***Description*:** Indicate if you have pimples or acne on your face or jawline (Yes/No).

11. ***Do you eat fast food regularly?***

- ***Label*:** FastFood

- ***Description***: Indicate if you consume fast food on a regular basis (Yes/No).

12. ***Do you exercise on a regular basis?***

- ***Label*:** Exercise

- ***Description*:** Indicate if you exercise regularly (Yes/No).

13. ***Have you been diagnosed with PCOS/PCOD?***

- ***Label*:** PCOS_PCOD

- ***Description*:** Indicate if you have been diagnosed with Polycystic Ovary Syndrome (PCOS) or Polycystic Ovary Disorder (PCOD) (Yes/No).

14. ***Do you experience mood swings?***

- ***Label*:** MoodSwings

- ***Description*:** Indicate if you experience mood swings (Yes/No).

15. ***Are your periods regular?***

- ***Label*:** RegularPeriods

- ***Description*:** Indicate if your periods are regular (Yes/No).

16. ***How long does your period last? (in Days, example- 1,2,3,4.....)***

- ***Label*:** PeriodDuration

- ***Description***: Enter the number of days your period typically lasts.

This dataset is intended for research purposes to improve understanding and management of women's health issues, providing valuable insights that can lead to better healthcare strategies and support.

Dataset Description For ChatBot

Menstrual Health Awareness Dataset: A Detailed Look

This dataset offers a valuable resource for training large language models (LLMs) to understand and respond effectively to topics related to menstruation. Here's a breakdown of its key aspects:

Context and Goal:

- Designed to train and improve the ability of LLMs to handle menstruation-related topics.
- Aims to equip LLMs with the capability to understand and generate accurate information about menstrual health, hygiene, and social issues surrounding it.

Dataset Structure:

- Two CSV files:
    - Training Data.csv: Contains 517 question-and-answer pairs for training the LLM.
    - Testing Data.csv: Includes 45 question-and-answer pairs for model validation and testing.

Data Content:

Covers a broad range of topics related to menstruation, including:

- Practices for menstrual health and hygiene
- Common menstrual disorders and their management

    - Addressing taboos and stigmas associated with menstruation
    - Nutritional considerations for menstrual health
    - General awareness and education about menstrual cycles

Dataset Description: English Menstrual Health and Hygiene Chatbot Conversations

This dataset was curated by Team Mai from Habib University, Pakistan, for their final year project. The project aims to develop a transformer-based chatbot that can understand and respond to English language queries related to menstrual health and hygiene.

Data Sources:

- Hello Clue
- Flo
- Always
- Kaggle
- Reddit
- Local data collected from stakeholders and doctors

Data Content:

The dataset consists of anonymized, multi-turn conversations focused on menstrual health and hygiene. It includes:

- **Intents:** These represent the user's main goal or question within the conversation (e.g., "sort", "information").
- **Text Format:** This indicates the type of text used by the user (e.g., "Hi", "Greeting").
- **Responses:** These are the chatbot's responses to user queries, providing accurate and informative answers.

Data Characteristics:

- **Conversational Flow:** The dataset prioritizes natural and engaging conversational flow, mimicking real-life interactions.
- **Information Accuracy:** The information provided by the chatbot is well-researched and aligns with best practices in menstrual health.

Detailed System Description for Menstrual Health Platform

Overview

This system is designed to provide a comprehensive platform for menstrual health education, support, and tools for women. The platform will include a blog, a period cycle calculator, a PCOS prediction tool, a support chatbot, a public forum, and interactive tutorials, all aimed at enhancing users' understanding and management of menstrual health.

Modules and Components

1. Blog Page

Content

The blog page will be a key component of the platform, offering a variety of content related to menstrual health:

- **Exploratory Data Analysis (EDA):**
  - **Purpose:** To present insightful data visualizations and global analyses on menstrual health trends.
  - **Features:** Interactive charts, graphs, and statistical summaries covering topics such as menstrual health issues, hygiene practices, and more.
- **Educational Articles:**
  - **Topics:** Menstrual health, hygiene practices, self-care during periods, wellness tips.
  - **Format:** Comprehensive articles, potentially with multimedia elements such as images and videos.
- **Guest Contributions:**
  - **Authors:** Healthcare professionals and guest bloggers.
  - **Content:** Diverse perspectives and expert advice on menstrual health.
- **Minimum Requirement:**
  - At least five comprehensive blog posts covering various aspects of menstrual health.

User Interface

- **Design:** Clean and intuitive layout with easy navigation.
- **Visual Aids:** Charts, graphs, and other visual aids to enhance the user experience and help users make informed decisions.

2. Period Cycle Calculator

Purpose

To help women predict their next menstrual period based on their cycle history.

Functionality

- **User Input:**
  - Date of last period.
  - Average cycle length.
- **Prediction:**
  - Calculates the next period date.
  - Provides insights into the user's menstrual cycle patterns.
- **Additional Features:**
  - Options for users to set notifications and reminders for the next period.
  - Educational information about menstrual cycles and what to expect.

3. PCOS Prediction Tool

Functionality

- **Questionnaire:**
  - Designed to assess the risk of Polycystic Ovary Syndrome (PCOS).
  - Covers symptoms, medical history, and lifestyle factors.
- **Results:**
  - Indicates the level of risk.
  - Provides suggestions for further medical consultation if needed.

Design

- **Confidentiality:** Ensure user privacy and confidentiality.
- **User Interface:** Simple and user-friendly design for easy navigation and use.

4. Chatbot for Support

Purpose

To provide instant support and information to users.

Capabilities

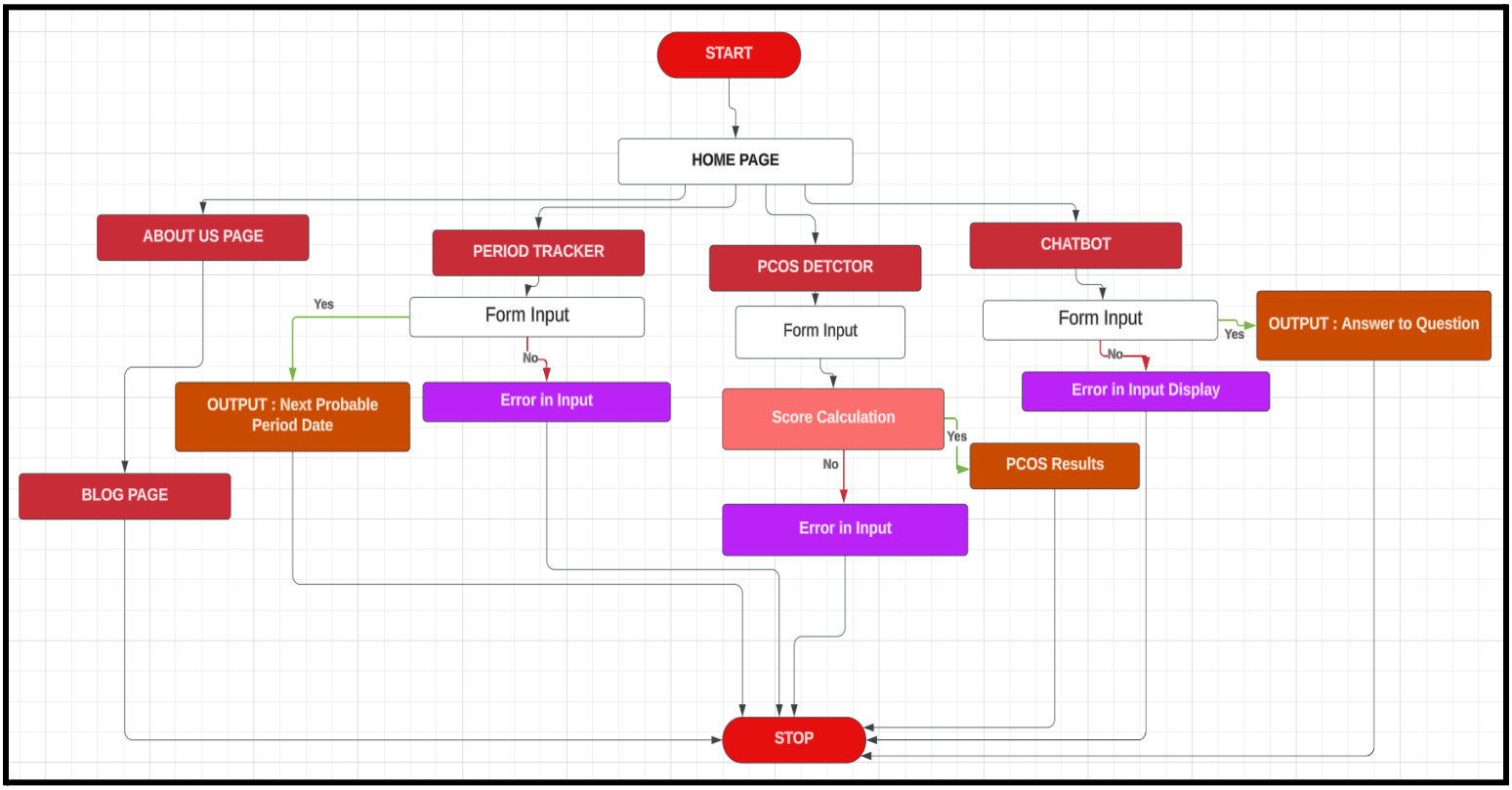
- **FAQ Responses:**
  - Answers frequently asked questions about menstrual health and hygiene.

- **Guidance:**
    - Directs users to relevant sections of the website for more information.
- **Emotional Support:**
    - Offers resources and support for users needing further assistance.

Availability

- **24/7 Availability:** Ensures users can get support at any time.
- **User Interface:** Designed to be friendly and easy to interact with.

Block Diagram



Roles And Responsibilities

Name	Registration No	Roles
Asmita Mondal	2348018	Frontend development, Backend integration, Chatbot development, Documentation
Swarnasish Banerjee	2348066	PCOD detection modelling and optimization, Development of deployable notebook, Background research
Sayan Pal	2348056	Blog development with EDA, Backend integration, Documentation, Period calculator development