**Menstrual Cycle Prediction System using Hybrid Machine Learning Algorithms: An Evaluation of Predictive Accuracy**

1\*Odirichukwu, J . C., 2Njoku, A.O., 3Odirichukwu, SP . C., 4Ndigwe, C, 5Nnoroka, J . U

1,2,5 Federal University of Technology, Owerri.

1\*chiomajaco6@gmail.com, jacinta.odirichukwu@futo.edu.ng;

2IT Health Data Officer/Monitoring and Evaluation Officer, Imo State Primary Health Development Agency. Owerri, Nigeria

4Senior Lecturer, Odumegwu Ojukwu University, Uli (OOU)

**Abstract**

A menstrual cycle prediction system for predicting a woman's ovulation day was developed in this study. To train the menstrual cycle datasets obtained from Kaggle, the proposed model used 22 powerful machine-learning algorithms. Twelve features were extracted from the datasets' 80 features and used for training to predict the Estimated Day of Ovulation. Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R2 score), Mean Absolute Percentage Error (MAPE), and Explained Variance Score (EVS) are the metrics used to determine the accuracy of the proposed system. However, when the R2 scores of the algorithms were considered, the Decision Tree algorithm had the highest R2 score of 0.9864. . The Decision Tree model's high R2 score indicates that it provides a very accurate and reliable prediction of ovulation day. The proposed system was built on Google Colab using modern programming languages like Python, the Django framework, HTML, and CSS, as well as development environments like VsCode and Jupyter Notebook.

**Keywords:** Ovulation length, luteal phase, Menstrual Cycle, and Machine Learning

1. **Introduction**

As Artificial Intelligence (AI)  technology advances, its impact is expected to spread across a wide range of industries, including healthcare, finance, transportation, and manufacturing. Nonetheless, concerns about the potential risks of AI remain, including job displacement, data breaches, and unintended consequences from autonomous decision-making [2]. However, family planning programs recognize the critical role of contraception access in promoting reproductive health outcomes, preventing unintended pregnancies, reducing maternal mortality and morbidity risks, and assisting individuals and couples in reaching their desired family size. Individuals have different contraceptive needs and preferences, which is acknowledged. As a result, family planning programs work hard to provide a wide range of contraceptive options, including hormonal methods (oral contraceptives, transdermal patches, vaginal rings, or injectables), barrier methods (e.g., condoms or diaphragms), long-acting reversible contraception (LARC) methods (intrauterine devices (IUDs) or contraceptive implants), and permanent methods (terilization procedures). Furthermore, these programs place an emphasis on providing comprehensive counseling and support services, such as accurate information dissemination about proper utilization techniques, potential side effects and associated risks, and effective management strategies for any difficulties encountered[1]. One of the major challenges in developing a Machine Learning algorithm for period tracking is the variability of menstrual cycles, which can be influenced by a variety of factors such as stress, weight changes, and certain medical issues [10].

Big data from a connected ovulation test was used to discover the degree of diversity in menstrual cycle length and ovulation day among women trying to conceive [3]. According to the current research review, future research in this field may focus on improving developed models by incorporating extra data sources, such as menstrual symptoms and hormonal data, and by conducting more extensive investigations to verify conclusions in the existing system. Based on historical data, this paper aims to predict the estimated ovulation day of women. Predicting ovulation day accurately is critical. It provides crucial information for making family planning decisions. With accurate knowledge of their estimated ovulation day, women can efficiently plan pregnancies, check the health of their reproductive systems, and take preventative measures to ensure overall health. The study encompasses the following objectives:

1. Evaluate existing AI-powered fertility prediction systems that are designed to accurately predict ovulation and fertile days while collecting extensive data from online repositories.
2. Create a real-time ensemble machine learning model for predicting ovulation  day in women.
3. Implement the proposed design using Python programming language, ensuring practical applicability.
4. Conduct a comparative analysis between the existing system and the newly developed system, employing standard performance metrics to assess their respective efficacy.
5. **Related Works**

The purpose of this paper was to explain how data from a self-tracking health app for female mobile phone users can be used to improve the accuracy of the next ovulation date prediction. We looked at the relationship between menstrual cycle duration, follicular phase length, and luteal phase length using data from 7043 women out of 8,000,000 users of a health care provider's mobile phone app who had reliable menstrual and ovulation records. Following that, we compared the linear function we devised to describe the relationship between ovulation timing and menstrual cycle length to the existing calendar-based approaches. There was a positive correlation between previous and upcoming menstrual cycle lengths, and the correlation between menstrual cycle length and follicular phase length was stronger than the correlation between menstrual cycle length and luteal phase length. The length of the follicular phase was found to be strongly positively correlated with the average length of previous cycles. The relationship between luteal phase lengths and mean cycle duration was also statistically significant. In the majority of cases, the optimized function method outperformed the Ogino method for forecasting the next ovulation day. Our method outperformed the ovulation date prediction method that uses the middle day of a typical menstrual cycle as the ovulation date. Because of the large number of patients, the authors were able to capture the correlations between the durations of the menstrual cycle, follicular phase, and luteal phase in greater detail than previous research [9].

Understanding your fertility and predicting your menstrual cycle are critical for better fecundity and health management. Previous research has used physiological indicators such as basal body temperature (BBT) and heart rate (HR) to forecast the fertile window and menses. Their precision, on the other hand, falls far short of expectations. Few researchers have looked into irregular menstruators. As a result, the goal of this paper was to develop algorithms capable of predicting the menstrual cycle and fertile window for both regular and irregular menstruators. The International Peace Maternity and Child Health Hospital in Shanghai, China, conducted this prospective observational cohort study. The study recruited participants from August to November 2020, and they were followed for at least four menstrual cycles.Participants wore the Huawei Band 5 to track their heart rate while an ear thermometer was used to measure BBT. Serum hormone levels and ovarian ultrasonography were used to determine the ovulation day. Women reported having periods on their own. We used linear mixed models to assess changes in physiological indicators. We also used machine learning to create probability function estimate models to forecast the fertile window and menstrual cycle. Data from 305 and 77 qualified cycles with confirmed ovulation from 89 regular menstruators and 25 irregular menstruators, respectively, were included in the study. BBT and HR were significantly higher during the fertile phase than the follicular phase and peaked during the luteal phase in regular menstruators (all P0.001). The physiological parameters of menstruating irregularly followed a similar pattern.We developed algorithms based on BBT and HR that predicted the fertile window with an accuracy of 87.46%, sensitivity of 69.30%, specificity of 92.00%, and AUC of 0.8993 and menses with an accuracy of 89.60%, sensitivity of 70.70%, specificity of 94.30%, and AUC of 0.7849 among regular menstruators. For irregular menstruators, the accuracy, sensitivity, specificity, and AUC for fertile window prediction were 72.51%, 21.00%, 82.90%, and 0.5808, respectively, and 75.90%, 36.30%, 84.40%, and 0.6759 for menses prediction [7].

Machine learning has the potential to improve the accuracy of period monitoring software by examining trends in data related to the menstrual cycle. It is critical for women's health to be able to forecast menstrual cycle timing in order to better understand and manage menstrual cycles, as well as to provide personalized reminders and notifications. One of the major challenges in developing a Machine Learning algorithm for period tracking is the variability of menstrual cycles, which can be influenced by a variety of factors such as stress, weight changes, and certain medical issues. Using time series forecasting algorithms such as ARIMA and STL, which are designed to estimate future values based on historical data, is one strategy. These algorithms can be trained with prior menstrual cycle data to forecast the timing of upcoming cycles. Another strategy is to use neural networks, which can simulate intricate patterns in data and can be used to forecast the start of the next period by examining patterns in information about the menstrual cycle, such as cycle length and symptoms. Random Forest and Gradient Boosting, ensemble methods used for classification and regression tasks, can be used to forecast the next period date by analyzing trends in menstrual cycle data, such as cycle duration and symptoms. Support vector machines (SVMs) is another machine learning approach for prediction, particularly in classification problems. Furthermore, by incorporating additional relevant data such as stress levels, weight fluctuations, and symptoms, the accuracy of predictions can be improved. Furthermore, machine learning can detect trends and patterns in data from multiple users, which may be useful in better understanding and controlling menstrual cycles.Machine learning has the potential to greatly improve the utility and accuracy of period monitoring apps, providing women with more tools for understanding and controlling their menstrual cycles [6].

1. described a non-invasive wearable fertility monitoring system and proposed a powerful and adaptable statistical learning method for identifying and forecasting ovulation based on the system's data. The device consists of an earpiece that measures ear canal temperature every five minutes while a person sleeps and a base station that sends the data to a smartphone app for analysis. Fit a Hidden Markov Model (HMM) with two hidden states of high and low temperature after establishing a data-cleaning methodology for data preprocessing to identify the most likely state of each time point using the anticipated probabilities. A post-processing method is developed for each participant that incorporates biorhythm data to generate a time-course biphasic profile. The performance of the suggested algorithms applied to data acquired by the device is compared to existing approaches in terms of matching rates with self-reported ovulation days verified by an ovulation test kit. An empirical study involving 34 users revealed significant improvements over conventional techniques in terms of detection accuracy (with sensitivity of 92.31%) and prediction power (23.07-31.55% greater). It was demonstrated that ovulation could be reliably predicted and identified using high-frequency temperature data collected by a non-invasive wearable device. Traditional methods of determining fertility are frequently untrustworthy or impractical. The wearable gadget and learning algorithm described in this study provide a user-friendly and dependable platform for tracking ovulation. This platform has the potential to have a significant impact on fertility research as well as practical family planning.

An online survey was used in this study to ask women about their actual experiences with period tracker apps, their attitudes toward using the apps, the information provided by the apps about ovulation, and how the app's accuracy in predicting period start dates affects their feelings and behaviors if their period arrives earlier or later than predicted. The online survey for this mixed-methods observational study included 50 multiple-choice and open-ended items. The survey was created and distributed via social media using Qualtrics. Anyone who had used a period tracker could take part. There were 330 complete responses out of a total of 375 responses, for an 88.0% completion rate. The respondents ranged in age from 14 to 54 years old, with a mean age of 26.0 (7.81). Knowing when I'm ovulating was chosen as the app's top feature by 29.7% (98/330) of respondents. When asked if their period had ever started earlier or later than the app predicted, 54.9% (189/330) said yes, while 72.1% (238/330) said no. Thematic analysis of early periods revealed four themes: feeling unaffected, frustrated/unprepared, anxious/stressed, and confused/intrigued. Participants were asked how they would feel if their period came earlier or later than expected. Period trackers, particularly for period due date and ovulation tracking, need to be more transparent about their intended purpose and accuracy. In many ways, qualitative research demonstrates the effects of incorrect forecasts on the health of users. This study encourages period tracker app developers to improve their products in order to educate their customers about their intended uses [5].

The majority of women of reproductive age require regular visits to a gynecologist or other doctor to assess menstrual health and fertility. Despite the fact that these assessments are critical for determining a person's reproductive health status, they frequently rely on memory-based self-reports, and the results are rarely evaluated at the population level. Menstrual tracking mobile apps have recently gained popularity. These apps allow us to evaluate the accuracy and tracking frequency of millions of self-observations, providing us with an unprecedented view of menstrual health and its evolution for large populations on a large scale and in great detail. The primary goal of this study was to describe app users' tracking behavior and general observation patterns to see if they were consistent with previous small-scale medical research. The secondary goal was to see if their precision allowed for ovulation timing detection and estimation, which is critical for reproductive and menstrual health. Two mobile apps used to implement the sympto-thermal fertility awareness approach collected retrospective self-observation data for 200,000 users, yielding a dataset of more than 30 million days of observations from more than 2.7 million cycles. According to the data analysis, up to 40% of the cycles in which users were looking for pregnancies had daily recordings. According to a modeling approach that used Hidden Markov Models to describe the data collected and estimate ovulation timing, only 24% of ovulations occurred between cycle days 14 and 15. In contrast, despite the fact that short luteal phases (10 days or less) were more frequently observed (in up to 20% of cycles), the average duration and range of the luteal phase were consistent with previous reports. The menstrual health topic, which has received very little research in the past, and its relationship to women's health in general, can be better understood using the digital epidemiology technique described here [8].

1. **Materials and Method**

Jupyter Notebook and Visual Studio code are the two main tools used in carrying out this research

* 1. **Methodology**

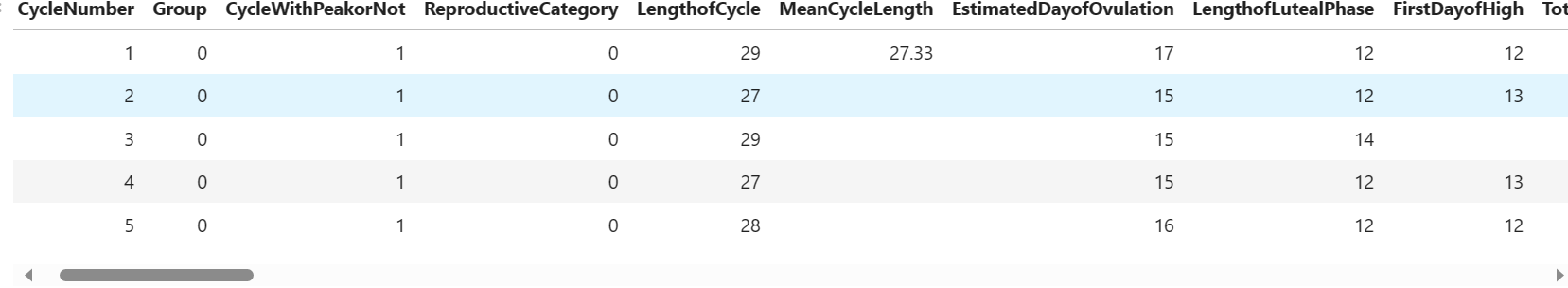
**Problem Definition**

The variability of menstrual cycles, which can be influenced by a range of factors including stress, changes in weight, and certain medical issues, is one of the major obstacles in designing a Machine Learning algorithm for period tracking. Women seeking to optimize conception require a precise and dependable fertility prediction system capable of identifying their most fertile days with precision. However, current systems rely on using variety of machine learning algorithms that fail to use individualized characteristics of each woman's menstrual cycle. The existing system data may not accurately reflect the general population because they came from fake data created by the research. The existing research suggested that future research in this field may concentrate on improving the models through the incorporation of extra data sources, such as menstrual symptoms and hormonal data, and through the execution of more extensive investigations to verify our conclusions.

**Data Collection**

Menstrual cycle dataset was collected from kaggle via this link <https://www.kaggle.com/datasets/nikitabisht/menstrual-cycle-data>

Sample Raw dataset before preprocessing



**Data Preprocessing**:

The dataset was preprocessed to remove irrelevant data, null values. Also, feature engineering was done on the dataset. Out of the 80 columns contained in the dataset, only 12 columns were extracted. These columns include:

**CycleNumber', 'LengthofCycle', 'LengthofLutealPhase', 'TotalNumberofHighDays', 'TotalNumberofPeakDays', 'UnusualBleeding', 'PhasesBleeding', 'IntercourseInFertileWindow', 'Age', 'BMI', 'Method', 'EstimatedDayofOvulation'**

The columns were then splitted into feature columns and target columns. The feature columns include: **‘CycleNumber', 'LengthofCycle', 'LengthofLutealPhase', 'TotalNumberofHighDays', 'TotalNumberofPeakDays', 'UnusualBleeding', 'PhasesBleeding', 'IntercourseInFertileWindow', 'Age', 'BMI', 'Method'**,

While the target column is **‘EstimatedDayofOvulation’**. The feature columns were used to predict the Ovulation day of a woman.

**Exploratory Data Analysis (EDA)**:

Here are the activities of the EDA of this research.

Jupyter notebook was used,

Basically these are the python libraries needed:

#Import the necessary librabries

import numpy as np

import pandas as pd

from sklearn.ensemble import RandomForestClassifier

from sklearn.tree import DecisionTreeClassifier

from sklearn.linear\_model import LogisticRegression

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import accuracy\_score, f1\_score, recall\_score, r2\_score

import matplotlib.pyplot as plt

import seaborn as sn

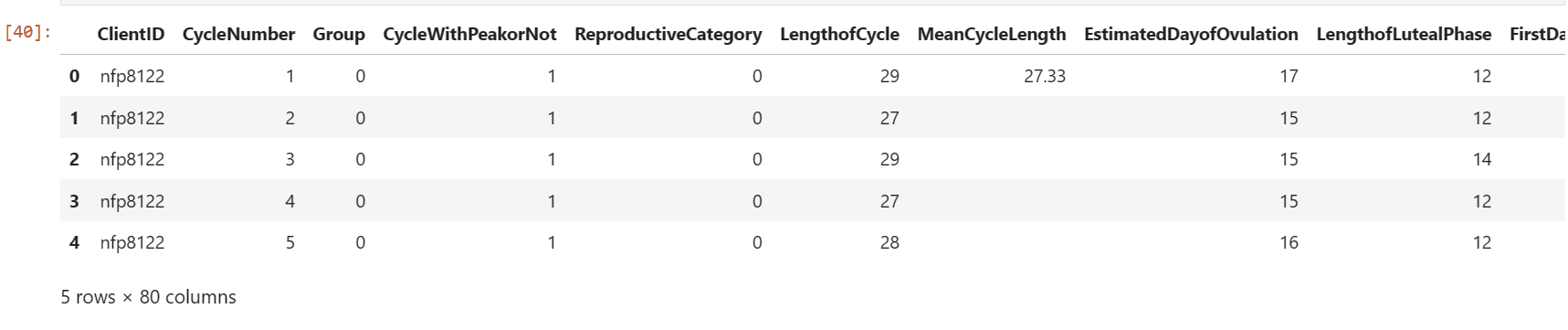
The dataset is loaded using this;

#Import the dataset

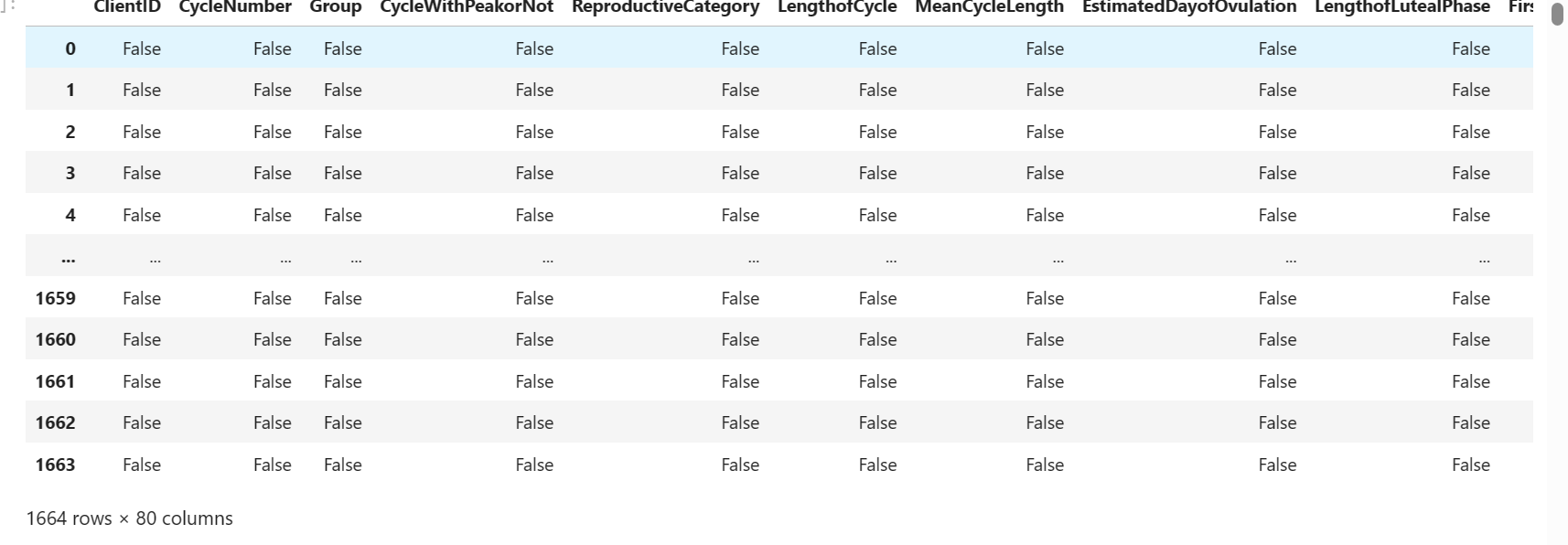
df = pd.read\_csv("C:/Users/hp/Desktop/MLs/DeployModelOvuLength/FedCycleData.csv")

#To view five rows out of the dataset

df.head()

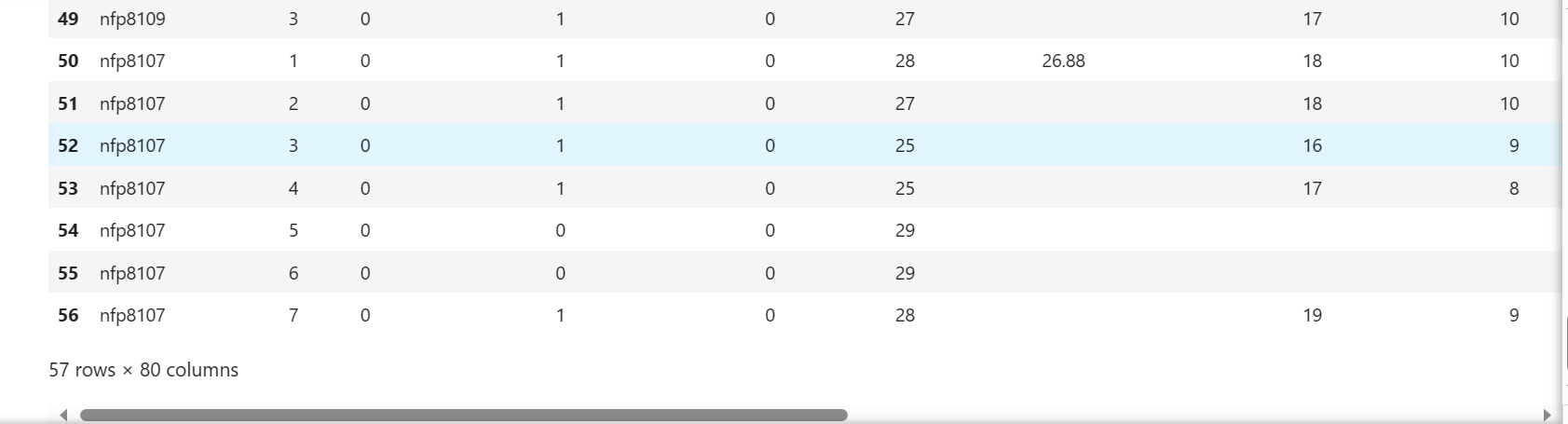


df.isnull()



# view first 57 rows

df.head(57)



The duplicates and nul values were all removed

**Feature Engineering**:

# Engineer relevant features from the data. This may involve selecting important features, creating new # ones, or transforming existing features to improve model performance. Here, relevant feature were

# selected to perform feature Engineering.

# Select relevant features and target variable (e.g., 'EstimatedDayofOvulation' LengthofLutealPhase,

# CycleNumber: To track the individual's cycle history.

# LengthofCycle: The length of the menstrual cycle.

# LengthofLutealPhase: The length of the luteal phase.

# TotalNumberofHighDays: The number of high fertility days.

# TotalNumberofPeakDays: The number of peak fertility days.

# UnusualBleeding: Information about unusual bleeding patterns.

# PhasesBleeding: Different phases of bleeding during the cycle.

# IntercourseInFertileWindow: Whether intercourse occurred during the fertile window.

# Age: Age of the individual.

# BMI: Body Mass Index.

# Method: Information about contraceptive methods used.

ovulation\_dataset = df[['CycleNumber', 'LengthofCycle', 'LengthofLutealPhase', 'TotalNumberofHighDays', 'TotalNumberofPeakDays', 'UnusualBleeding', 'PhasesBleeding', 'IntercourseInFertileWindow','Age','BMI','Method','EstimatedDayofOvulation']]

# 1. Replacing missing values(Nan) in CycleNumber with the mode

mode\_CycleNumber = ovulation\_dataset['CycleNumber'].mode()[0]

ovulation\_dataset['CycleNumber'].fillna(mode\_CycleNumber, inplace=True)

# 2. Replacing missing values(Nan) in LengthofCycle with the mode

mode\_LengthofCycle = ovulation\_dataset['LengthofCycle'].mode()[0]

ovulation\_dataset['LengthofCycle'].fillna(mode\_LengthofCycle, inplace=True)

# 3. Replacing missing values(Nan) in LengthofLutealPhase with the mode

mode\_LengthofLutealPhase = ovulation\_dataset['LengthofLutealPhase'].mode()[0]

ovulation\_dataset['LengthofLutealPhase'].fillna(mode\_LengthofLutealPhase, inplace=True)

# 4. Replacing missing values(Nan) in TotalNumberofHighDays with the mode

mode\_TotalNumberofHighDays = ovulation\_dataset['TotalNumberofHighDays'].mode()[0]

ovulation\_dataset['TotalNumberofHighDays'].fillna(mode\_TotalNumberofHighDays, inplace=True)

# 5. Replacing missing values(Nan) in TotalNumberofPeakDays with the mode

mode\_TotalNumberofPeakDays = ovulation\_dataset['TotalNumberofPeakDays'].mode()[0]

ovulation\_dataset['TotalNumberofPeakDays'].fillna(mode\_TotalNumberofPeakDays, inplace=True)

# 6. Replacing missing values(Nan) in UnusualBleedingwith the mode

mode\_UnusualBleeding = ovulation\_dataset['UnusualBleeding'].mode()[0]

ovulation\_dataset['UnusualBleeding'].fillna(mode\_UnusualBleeding, inplace=True)

# 7. Replacing missing values(Nan) in PhasesBleeding with the mode

mode\_PhasesBleeding = ovulation\_dataset['PhasesBleeding'].mode()[0]

ovulation\_dataset['PhasesBleeding'].fillna(mode\_PhasesBleeding , inplace=True)

# 8. Replacing missing values(Nan) in PhasesBleeding with the mode

mode\_IntercourseInFertileWindow = ovulation\_dataset['IntercourseInFertileWindow'].mode()[0]

ovulation\_dataset['IntercourseInFertileWindow'].fillna(mode\_IntercourseInFertileWindow, inplace=True)

# 9. Replacing missing values(Nan) in Age with the mode

mode\_Age = ovulation\_dataset['Age'].mode()[0]

ovulation\_dataset['Age'].fillna(mode\_Age , inplace=True)

# 10. Replacing missing values(Nan) in BMI with the mode

mode\_BMI = ovulation\_dataset['BMI'].mode()[0]

ovulation\_dataset['BMI'].fillna(mode\_BMI, inplace=True)

# 11. Replacing missing values(Nan) in Method with the mode

mode\_Method = ovulation\_dataset['Method'].mode()[0]

ovulation\_dataset['Method'].fillna(mode\_Method , inplace=True)

# 12. Replacing missing values(Nan) in EstimatedDayofOvulation with the mode

mode\_EstimatedDayofOvulation = df['EstimatedDayofOvulation'].mode()[0]

ovulation\_dataset['EstimatedDayofOvulation'].fillna(mode\_EstimatedDayofOvulation, inplace=True)

# Convert the column to int

ovulation\_dataset['CycleNumber'] = ovulation\_dataset['CycleNumber'].astype(int)

ovulation\_dataset['LengthofCycle'] = ovulation\_dataset['LengthofCycle'].astype(int)

ovulation\_dataset['LengthofLutealPhase'] = ovulation\_dataset['LengthofLutealPhase'].astype(int)

ovulation\_dataset['TotalNumberofHighDays'] = ovulation\_dataset['TotalNumberofHighDays'].astype(int)

ovulation\_dataset['TotalNumberofPeakDays'] = ovulation\_dataset['TotalNumberofPeakDays'].astype(int)

ovulation\_dataset['UnusualBleeding'] = ovulation\_dataset['UnusualBleeding'].astype(int)

ovulation\_dataset['PhasesBleeding'] = ovulation\_dataset['PhasesBleeding'].astype(int)

ovulation\_dataset['IntercourseInFertileWindow'] = ovulation\_dataset['IntercourseInFertileWindow'].astype(int)

ovulation\_dataset['Age'] = ovulation\_dataset['Age'].astype(int)

ovulation\_dataset['BMI'] = ovulation\_dataset['BMI'].astype(int)

ovulation\_dataset['Method'] = ovulation\_dataset['Method'].astype(int)

ovulation\_dataset['EstimatedDayofOvulation'] = ovulation\_dataset['EstimatedDayofOvulation'].astype(int)

#Splitting the dataset

'''After preprocessing the dataset, it was then splitted into features and target set.

X represent the features dataset, while Y represent the target dataset.

The drop() function used here is used to drop the 'EstimatedDayofOvulation' from the dataset so as to separate it from the features dataset.

It is therefore shown below:'''

X = ovulation\_dataset.drop('EstimatedDayofOvulation', axis= 1)

y = ovulation\_dataset['EstimatedDayofOvulation']

**Model Selection**

**Decision Tree algorithm, Random Forest, and Linear Regression and other 19 algorithms were chosen** as appropriate machine learning algorithms and models for th problem since this is a regression problem based on the dataset.

**Model Training**:

Train your selected models on your dataset. Use techniques like cross-validation to assess their performance and tune hyperparameters for better results.

In training the model, twenty two machine learning algorithm were used.

The libraries used are listed below:

****#Import the necessary librabries****

1. import numpy as np
2. import pandas as pd
3. from sklearn.ensemble import RandomForestClassifier
4. from sklearn.tree import DecisionTreeClassifier
5. from sklearn.linear\_model import LogisticRegression
6. from sklearn.model\_selection import train\_test\_split
7. from sklearn.metrics import accuracy\_score, f1\_score, recall\_score, r2\_score
8. import matplotlib.pyplot as plt
9. import seaborn as sn

**Model Evaluation:**

Evaluate the models using appropriate evaluation metrics (e.g., accuracy, F1-score, RMSE). Compare different models to select the best-performing one.

**Interpretability**:

If applicable, ensure that your model's decisions are interpretable and explainable. This is important for gaining insights and building trust in the model.

**Deployment**:

The Decision Tree Regressor model was deployed using python and django framework

**Documentation:**

The study was properly documented in the git hub repository as provided in the link

**Communication:**

The result is communicated properly in the result session using data visualization and tables

**Reproducibility:**

The code for this model could be found in github repository via this link <https://github.com/chiomajaco6/MenstrualCyclePrediction.> Researchers can follow the steps describe here to evaluate the results produced and to carry further research

**Iterate:**

Machine learning research is often an iterative process. Use your findings to refine your approach and experiment with different techniques to improve model performance.

**3.2 Analysis of the Existing System**

A correct menstrual cycle prediction is important for women's health since it enables people to take precautions to reduce discomforts related to cycles. Additionally, accurate prediction might be helpful for a woman's family planning and other significant life events. In this study, we investigated how to predict regular and irregular menstrual cycles using machine learning approaches. The AutoRegressive Integrated Moving Average, Huber Regression, Lasso Regression, Orthogonal Matching Pursuit, and Long Short-Term Memory Network are some of the time series forecasting algorithm approaches that we used. In addition, we created fake data to serve our objectives. The findings demonstrated that machine learning techniques can be used to precisely forecast the start and length of menstrual periods.

In this study, we used machine learning techniques to forecast the menstrual cycle utilizing information from a given model's generated data. The algorithms can spot patterns and correlations that humans might not see right away. This may result in fresh perspectives and understandings in the study of menstrual cycle prediction. Furthermore, our findings imply that machine learning models are capable of making reliable, low-error predictions of menstrual cycle phase. Our research has significant health implications for women and may be used to guide individual reproductive health choices including family planning and fertility therapy. Additionally, the algorithms can be trained on unique data, making it possible to forecast menstrual cycle trends specifically for each user. Since conventional prediction techniques may not be as effective for women with irregular periods, this could be very helpful to them. Future research in this field may concentrate on improving our models through the incorporation of extra data sources, such as menstrual symptoms and hormonal data, and through the execution of more extensive investigations to verify our conclusions.

The following are drawbacks of the existing system:

Since conventional prediction techniques may not be as effective for women with irregular periods, this could be very helpful to them. Future research in this field may concentrate on improving our models through the incorporation of extra data sources, such as menstrual symptoms and hormonal data, and through the execution of more extensive investigations to verify our conclusions.

[11].

**3.3 Design of the New System**

The new system will enhance the existing system by implementing a user-friendly interface where users can be able to predict their ovulation day. Twenty two different machine learning models were built to perform the prediction. Among the twenty two models, Decision tree model has the highest R2 score of 0.9864 and was used as the best model for the prediction. Also, the model was deployed on a web application using Django.

The introduction of the new system signifies a notable advancement over the existing framework, characterized by the introduction of a user-friendly interface that offers a seamless experience for users to predict their menstrual cycle length accurately. This innovation addresses a crucial need in women's reproductive health, providing them with a practical tool for family planning and personal well-being.

The development process entailed the creation of three distinct machine learning models, each specifically designed for accurate prediction: linear regression, Random Forest, and Decision Tree. These models, driven by advanced algorithmic methodologies, constitute the core of the prediction mechanism. Their diversity caters to a range of data patterns and ensures the highest predictive accuracy attainable within the system.

The performance evaluation of these models is of utmost importance. The Decision Tree model has the highest accuracy among them, exhibiting the highest R2 score of 0.9864. The R2 score provides insights into how well the predictions made by the system align with the actual observed outcomes. In this context, the Decision Tree model's R2 score signifies its ability to accurately capture and predict variations in menstrual cycle length based on the selected features.

This exceptional R2 score of 0.9864 positions the Decision Tree model as the optimal choice for menstrual cycle length prediction, underscoring its superior predictive capabilities over the other models. This strategic selection is informed by the model's ability to discern complex patterns within the dataset, resulting in robust and dependable predictions.

Furthermore, the Decision Tree model's seamless integration into a web application environment adds an extra layer of accessibility and usability. This is achieved by utilizing the Django framework, a versatile and efficient web development tool. Deploying the model within a web application broadens its reach and usability, allowing users to easily interact with it and obtain real-time predictions. The interface has 11 input fields that represent the predicting variables. The user enters information about the input data and presses the predict button to learn about the estimated ovulation day.

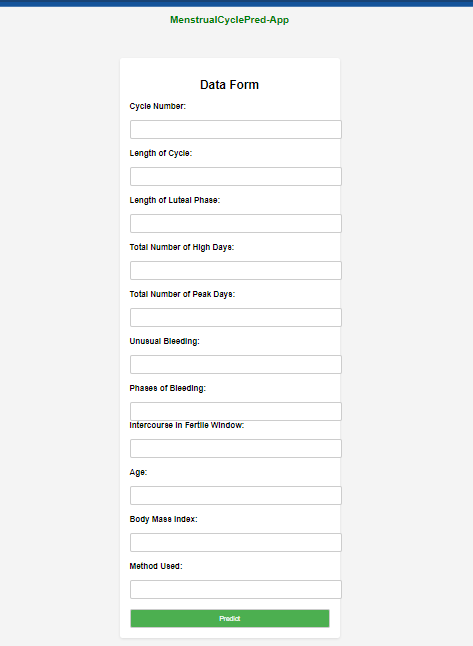


Figure 1: Home page of the Prediction System

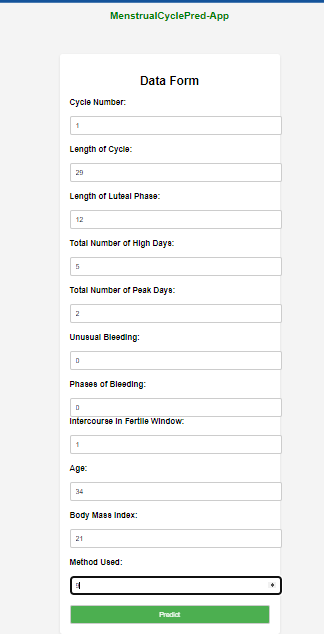
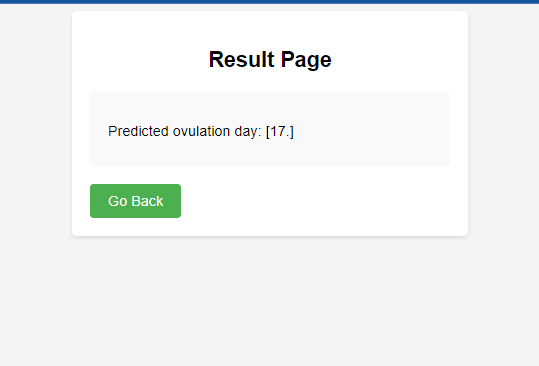


Figure 2: Making prediction with the proposed system



**Figure 3: Result of the prediction**

**Results**

Table 1: Proposed System Results Using only Eleven Featuresdf[['CycleNumber', 'LengthofCycle', 'LengthofLutealPhase', 'TotalNumberofHighDays', 'TotalNumberofPeakDays', 'UnusualBleeding', 'PhasesBleeding', 'IntercourseInFertileWindow','Age','BMI','Method','EstimatedDayofOvulation']]

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SN** | **MODEL** | **MSE**  **(Mean Squared Error)** | **RMSE**  **(Root Mean Squared Error)** | **MAE**  **(Mean Absolute Error)** | **R-squared** | **MAPE**  **(Mean Absolute Percentage Error)** | **EVS**  **(Explained Variance Score)** | **MAE**  **(Median Abslute Error)** |
| 1 | Decision Tree Regressor | 0.45081 | 0.67143 | 0.18852 | 0.98648 | 0.010979 | 0.98651 | 0.0 |
| 2 | Random Forest | 3.30327 | 1.81749 | 0.86065 | 0.90094 | 0.05052 | 0.90401 | 0.0 |
| 3 | Linear Regression | 26.70387 | 5.16757 | 4.25094 | 0.19926 | 2056701047386108 | 0.42958 | 3.90820 |
| 4 | Huber Regression | 25.01530 | 5.00152 | 4.08101 | 0.24989 | 2229126066185374.5 | 0.41314 | 3.79004 |
| 5 | Least Angle Regression | 26.70387 | 5.16757 | 4.25094 | 0.19926 | 2056701047386099.5 | 0.42958 | 3.90820 |
| 6 | Lasso Regresion | 25.06131 | 5.1675 | 4.05318 | 0.24851 | 2265472711515430.5 | 0.42399 | 3.72398 |
| 7 | Lasso Least Angle Regression | 25.06128 | 5.00612 | 4.05317 | 0.24851 | 2265461034058648.0 | 0.42399 | 3.72398 |
| 8 | **\***Dummy Regressor | 33.43917 | 5.78266 | 3.94681 | -0.00270 | 5619849841869245.0 | 0.0 | 2.49891 |
| 9 | Ridge Regressor | 26.68487 | 5.16574 | 4.24981 | 0.19983 | 2055881142848087.8 | 0.43000 | 3.90754 |
| 10 | K-Nearest Neighbour | 5.65819 | 2.37869 | 1.28442 | 0.8303 | 339615709604988.25 | 0.83876 | 0.80000 |
| 11 | Elastic Net | 24.97420 | 4.99741 | 4.05666 | 0.25112 | 2218799889360717.8 | 0.42720 | 3.71100 |
| 12 | Orthogonal Matching Pursuit | 22.68885 | 4.76328 | 3.60559 | 0.31965 | 2698083901957060.5 | 0.32227 | 2.62806 |
| 13 | LightGradient Boosting Matching | 0.6301 | 0.79382 | 0.34537 | 0.98110 | 218959992899727.28 | 0.981314 | 0.116121 |
| 14 | Bayesian Ridge | 26.34548 | 5.13278 | 4.22268 | 0.2100 | 218959992899727.28 | 0.43397 | 3.89005 |
| 15 | Passive Aggressive Regression | 19.94411 | 4.46588 | 3.39856 | 0.401957 | 1960971980850723.2 | 0.40526 | 2.62654 |
| 16 | Catboost | 0.73534 | 0.85752 | 1.05825 | 0.97795 | 355703903252822.7 | 0.977982 | 0.0962 |
| 17 | Gradient boosting | 0.74677 | 0.86416 | 0.28629 | 0.97760 | 265606557593265.97 | 0.97795 | 0.07067 |
| 18 | Extra Tree Regresor | 8.42755 | 2.90302 | 2.13404 | 0.74729 | 633743464586532.5 | 0.755602 | 1.55512 |
| 19 | Xgb | 0.52388 | 0.72380 | 0.27108 | 0.98429 | 80936650934372.72 | 0.98437 | 0.08360 |
| 20 | Ada Boost Regressor | 2.72667 | 1.65126 | 1.05825 | 0.40195 | 0.06624 | 0.91835 | 0.73964 |
| 21 | ARIMA | 37.49908 | 6.12364 | 5.49874 | -1.9800 | 0.39423 | 3.21113 | 5.95435 |
| 22 | LSTM | 0.023616 | 0.15367 | 0.11361 | 0.14613 | 0.27466 | 0.14722 | 0.08839 |

Table 2: Result of the New System

|  |  |  |
| --- | --- | --- |
| Sn. | Evaluated parameters | Values |
| 1 | Software development model | Structured Systems Analysis and Design Methodology |
| 2 | Number of application | 1 (Web) |
| 3 | Development tool | HTML, CSS, Django, Python, Jupyter Notebook. |
| 4 | Machine Learning Models | Decision Tree Regression, Random Forest, Logistic Regression, etc. |
| 5 | Best fit Machine Learning Model. | Decision Tree Regressor |
| 7 | R2 value of the best fit Machine Learning Model | 0.9864 |

Table 3: Comparative analysis of the Existing and New Systems

|  |  |  |  |
| --- | --- | --- | --- |
| Sn. | Parameters for comparison | VALUES  Existing New  system System | |
| 1 | Software development model | Object Oriented Analysis and Design Methodology (Agile model) | Structured Systems Analysis and Design Methodology |
| 2 | Number of applications | None | 1 (Web application) |
| 3 | Development tool | Python | HTML, CSS, Django, Python, Jupyter Notebook. |
| 4 | Machine Learning Models | AutoRegressive Integrated Moving Average, Huber Regression, Lasso Regression, Orthogonal Matching Pursuit, and Long Short-Term Memory Network. | Logistic Regression, Decision tree Regression, and Random Forest Regression, etc. |
| 5 | R2 Score | Not specified | 0.9864 |
| 6 | Best Machine Learning Model | Not specified | Decision Tree regression |

**Discussion**

This section presents the results and discuss the findings of the research project focused on developing a new system for ovulation day prediction. The system utilizes AI-powered algorithms, including 22 algorithms to accurately forecast the estimated ovulation day of women. Figure 1 displays the architectural design of the new system. It showcases the components and their interactions, highlighting the required data that are used for predicting the cycle length. Figure 2 process taken by the user to make prediction by entering the required variables for prediction. This user-friendly interface allows women to receive prompt predictions of their estimated ovulation day, empowering them with valuable information for family planning and reproductive health management. Figure 3 represents the results page of the prediction. Table 1 showcases accuracy parameters for twenty two models. The parameters include metrics such as Mean Square Error (MSE), Root Mean Squared Error (RMSR), Mean Absolute Error (MAE), R-Squared (RS), Mean Absolute Percentage Error (MAPE), Explained Variance Score (EVS), Median Absolute Error (MAE). These metrics helps asses the performance and effectiveness of the models. The R2 score measures the proportion of the variance in the ovulation day that can be explained by the models. A higher R2 score indicates a better fit of the mode to the data. Table 2 presents the results of the new system showing the parameters for evaluation. Table 3 offers a comparative analysis of the existing system and the new system developed in this research project. In conclusion, the results demonstrate the successful development and implementation of the new system for ovulation day prediction. The new system leverages AI-algorithms and provides accurate forecasts, empowering women with valuable information for their reproductive health management. The comparative analysis highlights the superiority of the new system over the existing one, emphasizing its enhanced accuracy and usability. These findings contribute to the field of women’s health and provide a valuable tool for family planning and fertility management.

**Contribution to Knowledge**

The contribution to knowledge of this ovulation day prediction app with an R2 score of 0.9864 lies in its ability to provide a predictive model that can estimate the estimated ovulation day of a woman. This contributes to women’s reproductive health by offering insights into their ovulation patterns and helping them plan their daily activities, anticipate the timing of their periods, and track their fertility. The R2 score of 0.9864 indicates that the developed prediction model explains 98.64% of the variance in the estimated ovulation day based on the input variables . This implies that the model has high predictive power, capturing a large portion of the variability in ovulation day based on the available data. By deploying this ovulation day prediction app using Django, the contribution to knowledge extends to the practical implementation of the predictive model in a web application accessible to users. The web app provides a user-friendly interface for entering the required input variables and obtaining the estimated ovulation day predictions. This deployment enables widespread access to the prediction model, empowering women to make informed decisions about their menstrual health. Furthermore, the deployment of the app using Django leverages the web development frameworks features, such as URL routing, form handling, and rendering templates, to ensure a seamless user experience. Django’s robustness and scalability make it suitable for handling user interactions, managing data etc.

**Future Work**

Based on the limitations of the study, further research should expand the scope of the study to cover the development of an internet-free mobile application.

**Conclusion**

Finally, the researchers developed and deployed a web application for an AI-powered ovulation day prediction system using Django. The study's goal was to use twenty-two algorithms to accurately predict women's estimated ovulation day. The Decision Tree Regressor algorithm outperformed the others, achieving the highest R2 score after thorough analysis and evaluation. We made fertility prediction capabilities easily accessible to women by deploying this system as a web application. The app has a user-friendly interface where women can enter relevant data such as cycle number, cycle length, luteal phase length, and so on to get accurate predictions of their estimated ovulation day. The successful deployment of the system on a web application built with Django opens the possibility to future enhancements and developments. It can be enhanced with features such as personalized insights, fertility tracking, and reminders. User feedback and engagement will be critical for the system's continuous improvement and refinement. The AI-powered fertility prediction system, which was built as a Django web application, represents a significant advancement in reproductive health technology. It gives women a simple and accurate way to predict their ovulation day, allowing them to take control of their fertility and make informed decisions about their reproductive health. The app will allow women to track their expected ovulation day over time. They can gain insight into their menstrual patterns, identify any irregularities or abnormalities, and plan for future menstrual cycles using this data. Knowing one's expected ovulation day is critical for fertility awareness and family planning. In addition, the ovulation day can be a useful tool for tracking overall reproductive health.

**Reference**

1. Solo, J., & Festin, M. (2019). Provider bias in family planning services: A review of its meaning and manifestations. Global Health Science and Practice.

2. Davidson, L., & Boland, M. R. (2020). Enabling pregnant women and their physicians to make informed medication decisions using artificial intelligence. Journal of Pharmacokinetics and Pharmacodynamics, 47, 305-318.

3. Soumpasis, I., Grace, B., & Johnson, S. (2020). Real-life insights on menstrual cycles and ovulation using big data. Human Reproduction Open, 2020(2), hoaa011.

4. Luo, L., She, X., Cao, J., Zhang, Y., Li, Y., & Song, P. X. (2019). Detection and prediction of ovulation from body temperature measured by an in-ear wearable thermometer. IEEE Transactions on Biomedical Engineering, 67(2), 512-522.

5. Broad, A., Biswakarma, R., & Harper, J. C. (2022). A survey of women’s experiences of using period tracker applications: Attitudes, ovulation prediction and how the accuracy of the app in predicting period start dates affects their feelings and behaviours. Women's Health, 18, 17455057221095246.

6. Morini, D., Melli, B., Spaggiari, G., Furini, C., Nicoli, A., Valli, B., ... & Villani, M. T. (2022). P-585 The (decision) tree of fertility: an innovative decision-making algorithm in assisted reproduction technique. Human Reproduction, 37(Supplement\_1), deac107- 539.

7. Yu, J. L., Su, Y. F., Zhang, C., Jin, L., Lin, X. H., Chen, L. T., ... & Wu, Y. T. (2022). Tracking of menstrual cycles and prediction of the fertile window via measurements of basal body temperature and heart rate as well as machine-learning algorithms. Reproductive Biology and Endocrinology, 20(1), 1-12.

8. Symul, L., Wac, K., Hillard, P., & Salathé, M. (2019). Assessment of menstrual health status and evolution through mobile apps for fertility awareness. NPJ digital medicine, 2(1), 64.

9. Sohda, S., Suzuki, K., & Igari, I. (2017). Relationship between the menstrual cycle and timing of ovulation revealed by new protocols: analysis of data from a self-tracking health app. Journal of medical Internet research, 19(11), e391.

10. Thakur, T., Kadam, S., Patil, N., & Achrekar, C. (2023). Machine Learning in Period, Fertility and Ovulation Tracking Application.

11. Rego, R. C. (2023). Predictive Modeling of Menstrual Cycle Length: A Time Series Forecasting Approach. arXiv preprint arXiv:2308.07927.