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

1\*Odirichukwu, J . C., 2Odirichukwu, SP . C., 3Ndigwe, C,

1 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

3Senior Lecturer, Odumegwu Ojukwu University, Uli (OOU)

**Abstract**

Accurate fertility prediction holds great importance for women, providing vital information for well-informed family planning choices. With precise knowledge of their menstrual cycle lengths, women can effectively plan pregnancies, monitor their reproductive well-being, and take proactive steps to ensure overall health. Although there have been efforts made towards accurate fertility prediction, lapses exist in terms of accuracy and efficiency. This paper developed a real-time hybrid model for an AI-powered fertility prediction system catering to women's reproductive health. The proposed model model three powerful machine learning algorithms: Decision Tree, Random Forest, and Linear Regression, with the aim of accurately forecasting the menstrual cycle length of women. To validate and assess the model's performance, extensive experimentation is conducted on a diverse dataset comprising real-world reproductive health parameters. The metric used to gauge the accuracy of the new system is the R2 score which is a crucial evaluation metric used to assess the predictive accuracy of a system. The system produces an R2 score of 0.58, in contrast to the existing system's R2 score of 0.37 thus, establishing itself as a veritable alternative. The system design was implemented using modern programming languages such as Python, Django framework, HTML, CSS, along with development environments like VsCode and Jupyter Notebook.

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

1. **Introduction**

The progression of artificial intelligence (AI) has been fueled by advancements in computer technology, such as enhanced computational power and the capacity to handle extensive volumes of data. Artificial Intelligent (AI) algorithms are constructed using mathematical models, enabling agents to extract patterns from data and make predictions based on acquired knowledge. Numerous approaches exist for constructing AI systems, including rule-based systems, neural networks, genetic algorithms, and fuzzy logic, each with its own strengths and weaknesses. Researchers are actively exploring novel methods and techniques to enhance AI performance. Robotics stands as a primary field for applying AI, where intelligent agents govern the movements and actions of machines. Additionally, AI finds applications in natural language processing, image recognition, and predictive analytics. As AI technology progresses, its influence is expected to permeate numerous industries, encompassing healthcare, finance, transportation, and manufacturing. Nevertheless, concerns regarding potential risks associated with AI persist, including job displacement, privacy breaches, and unintended consequences stemming from autonomous decision-making [2]. Family planning programs recognize the essential role of contraception accessibility in promoting reproductive health outcomes, averting unintended pregnancies, mitigating maternal mortality and morbidity risks, and facilitating individuals and couples in achieving their desired family size. However, it is acknowledged that diverse contraceptive needs and preferences exist among individuals.Consequently, family planning programs strive to provide a comprehensive array of contraceptive options, encompassing hormonal modalities (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). Moreover, these programs emphasize the delivery of comprehensive counseling and support services, including accurate information dissemination regarding proper utilization techniques, potential side effects and associated risks, and effective management strategies for any encountered difficulties [1]. 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 [10].

The degree of diversity in menstrual cycle length and ovulation day among women seeking to conceive was discovered using big data from a connected ovulation test [3].

The existing research, as the current research review suggests, suggests future research in this field may concentrate on improving developed 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 conclusions in the existing system.

This paper aims at predicting menstrual cycle length of women based past historical data. Accurate fertility prediction is crucial. It provides important knowledge for making decisions about family planning. Women can efficiently plan pregnancies, check the health of their reproductive systems, take preventative measures to guarantee overall health with precise knowledge of lengths of their menstrual cycles.

The study encompasses the following objectives:

Evaluate existing AI-powered fertility prediction systems designed for accurately predicting ovulation and fertile days, while gathering comprehensive data from online repositories.

Devise a real-time ensemble machine learning model specifically tailored for predicting ovulation and fertile days in women.

Enhance system performance through the utilization of a hybrid approach incorporating linear regression, machine learning, and artificial neural network algorithms.

Implement the proposed design using Python programming language, ensuring practical applicability.

Conduct a comparative analysis between the existing system and the newly developed system, employing standard performance metrics to assess their respective efficacy.

1. **Related Works**

This paper aimed to explain how the information from a self-tracking health app for female mobile phone users can be used to increase the precision of the next ovulation date prediction. We examined the association between the menstrual cycle duration, follicular phase length, and luteal phase length using the data of 7043 women out of 8,000,000 users of a mobile phone app of a health care provider who had reliable menstrual and ovulation records. After that, we compared the linear function we had fitted to the link between ovulation timing and menstrual cycle length to the existing calendar-based approaches. There was a positive correlation between the lengths of previous and upcoming menstrual cycles, and the correlation between the length of the menstrual cycle and the length of the follicular phase was stronger than the correlation between the length of the menstrual cycle and the length of the luteal phase. The length of the follicular phase and the mean length of previous cycles were found to be strongly positively correlated. Additionally statistically significant was the link between the lengths of the luteal phase and the mean cycle duration. The method based on the optimized function, beat the Ogino method for forecasting the following ovulation day in the majority of the individuals. The ovulation date prediction method that uses the middle day of a typical menstrual cycle as the ovulation date was likewise outperformed by our approach. The authors were able to capture the correlations between the durations of the menstrual cycle, follicular phase, and luteal phase in more detail than earlier research due to the huge number of patients [9].

For better fecundity and health management, understanding of your fertility and menstrual cycle prediction are crucial. Previous research has employed physiological indicators to forecast the fertile window and menses, such as basal body temperature (BBT) and heart rate (HR). Their precision, however, falls far short of expectations. Few scholars have also studied irregular menstruators. Therefore, the goal of this paper was to create algorithms that could predict the menstrual cycle and the fertile window for both regular and irregular menstruators. This prospective observational cohort study was carried out at the Shanghai, China's International Peace Maternity and Child Health Hospital. Recruitment for the study took place from August 2020 to November 2020, and participants were monitored for at least four menstrual cycles. Participants wore the Huawei Band 5 to track their heart rate while measuring BBT with an ear thermometer. The ovulation day was identified using serum hormone levels and ovarian ultrasonography. Women self-reported having periods. In order to evaluate changes in physiological indicators, we employed linear mixed models. We also created probability function estimate models using machine learning to forecast the fertile window and menstrual cycle. We included data from 305 and 77 qualified cycles with confirmed ovulation from 89 regular menstruators and 25 irregular menstruators, respectively. For regular menstruators, BBT and HR were signifcantly higher during fertile phase than follicular phase and peaked in the luteal phase (all P<0.001). The physiological parameters of irregular menstruators followed a similar trend. Based on BBT and HR, we developed algorithms that predicted the fertile window with an accuracy of 87.46%, sensitivity of 69.30%, specifcity of 92.00%, and AUC of 0.8993 and menses with an accuracy of 89.60%, sensitivity of 70.70%, and specifcity of 94.30%, and AUC of 0.7849 among regular menstruators. For irregular menstruators, the accuracy, sensitivity, specifcity and AUC were 72.51%, 21.00%, 82.90%, and 0.5808 respectively, for fertile window prediction and 75.90%, 36.30%, 84.40%, and 0.6759 for menses prediction [7]. By examining trends in the data related to the menstrual cycle, machine learning has the potential to enhance the accuracy of period monitoring software. In order to better understand and manage menstrual cycles, as well as to provide individualized reminders and notifications, it is crucial for women's health to be able to forecast the time of menstrual cycles. 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. Using time series forecasting algorithms like ARIMA and STL, which are intended to estimate future values based on historical data, is one strategy. To forecast the timing of upcoming cycles, these algorithms can be trained using prior menstrual cycle data. Another strategy is to employ neural networks, which can simulate intricate patterns in data and be used to forecast the start of the following period by examining patterns in information about the menstrual cycle, such as cycle length and symptoms. By examining trends in menstrual cycle data, such as cycle duration and symptoms, Random Forest and Gradient Boosting, ensemble methods used for classification and regression tasks, can be utilized to forecast the next period date. Another machine learning approach for prediction, specifically in classification issues, is support vector machines (SVMs). Additionally, by integrating additional pertinent data like details on stress levels, weight fluctuations, and symptoms, predictions' accuracy can be increased. Additionally, machine learning can spot trends and patterns in data from various users that could be helpful in better comprehending and controlling menstrual cycles.

Machine learning has the potential to significantly increase the utility and accuracy of period monitoring apps, giving women greater tools for comprehending and controlling their menstrual cycles [6]. [4] introduced a non-invasive wearable fertility monitoring system and suggest a powerful and adaptable statistical learning method to identify and forecast ovulation based on the data collected by the system. The device is made up of an earpiece that takes ear canal temperature readings every five minutes while a person is sleeping at night and a base station that sends the information to a smartphone app for analysis. To identify the most likely state of each time point using the anticipated probabilities, fit a Hidden Markov Model (HMM) with two hidden states of high and low temperature after establishing a data-cleaning methodology for data preprocessing. For each participant, a post-processing method is created that incorporates biorhythm data to create a time-course biphasic profile. In terms of matching rates with self-reported ovulation days verified by an ovulation test kit, the performance of the suggested algorithms applied to data acquired by the device is compared with existing approaches. A group of 34 users participated in an empirical investigation, and the results showed considerable advancements over the conventional techniques in terms of detection accuracy (with sensitivity of 92.31%) and prediction power (23.07-31.55% greater). It was showed that it was possible to anticipate and identify ovulation reliably using high-frequency temperature data gathered by a non-invasive wearable device. Conventional methods for gauging fertility are frequently unreliable or impractical. A user-friendly and reliable platform for tracking ovulation is provided by the wearable gadget and learning algorithm reported in this study. This platform may have a significant impact on both fertility research and practical family planning.

This study used an online survey to ask women about their actual experiences using period tracker apps, their attitudes toward using the apps, the information the apps provided regarding ovulation, and how the app's accuracy in predicting period start dates affects their feelings and behaviors if their period comes earlier or later than predicted. 50 multiple-choice and open-ended items were included in the online survey for this mixed-methods observational study. Utilizing Qualtrics, the survey was created and distributed over social media. Anyone who had used a period tracker was eligible to participate. 330 complete responses were gathered out of a total of 375 responses, for a completion percentage of 88.0%. The respondents' mean age was 26.0 (7.81) and they ranged in age from 14 to 54. To know when I'm ovulating was chosen as the top feature of the app by 29.7% (98/330) of respondents. When asked if their period had ever started earlier or later than the app had anticipated, 54.9% (189/330) of respondents indicated it had, while 72.1% (238/330) said it had. Thematic analysis of periods starting early revealed four themes: feeling unaffected, being frustrated/unprepared, feeling anxious/stressed, and feeling confused/intrigued. Participants were asked how they felt if their period arrived earlier or later than planned. Particularly for period due date and ovulation tracking, period trackers need to be more transparent about their intended purpose and accuracy. Qualitative investigation demonstrates the effects of erroneous forecasts on the users' health in many ways. This study urges period tracker app developers to upgrade their products to make their consumers aware of their intended uses [5].

Regular appointments to a gynecologist or other doctor are usually necessary for the majority of women of reproductive age in order to assess menstrual health and fertility. Despite the fact that these assessments are crucial for learning about a person's reproductive health status, they frequently rely on memory-based self-reports, and the outcomes are hardly ever evaluated at the population level. Mobile apps for menstrual tracking have gained a lot of traction recently. These apps enable us to assess the accuracy and tracking frequency of millions of self-observations, giving us an unmatched view of menstrual health and its evolution for large populations on a scale and in great detail. This study's main objective was to describe app users' tracking behavior and general observation patterns in order to determine whether or not they were consistent with earlier small-scale medical research. The secondary goal was to find out if their accuracy enabled ovulation timing detection and estimation, which is crucial for reproductive and menstrual health. Two mobile apps used to implement the sympto-thermal fertility awareness approach provided retrospective self-observation data, resulting in a dataset of more than 30 million days of observations from more than 2.7 million cycles for 200,000 users. The data analysis revealed that up to 40% of the cycles in which users were looking for pregnancies had daily recordings. Only 24% of ovulations occurred between cycle days 14 and 15, according to a modeling approach that used Hidden Markov Models to describe the data collected and estimate ovulation timing. In contrast, the average duration and range of the luteal phase were in line with previous reports, despite the fact that short luteal phases (10 days or less) were more frequently observed (in up to 20% of cycles). The menstrual health topic, which has previously received very little research, and its relationship to women's health in general can be better understood with the use of the digital epidemiology technique given here [8].

1. **Materials and Method**

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

* 1. **Methodology**

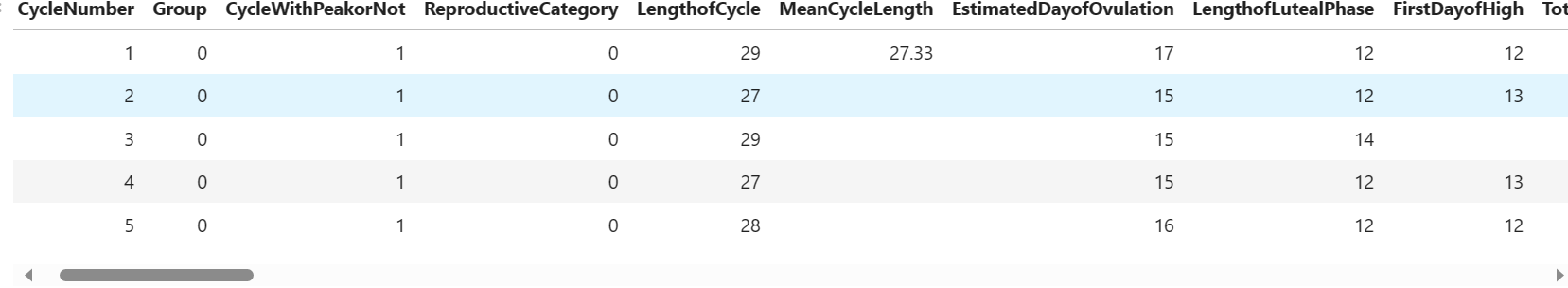
**Problem Definition**

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 Ovulution day of a woman.

The menstrual cycle length refers to the duration between the start of one menstrual period (also known as menstruation or the onset of bleeding) and the start of the next menstrual period. It represents the time it takes for a complete cycle of hormonal and physiological changes to occur within a woman’s reproductive system. The length of the menstrual cycle can vary among individuals and may also vary from cycle to cycle for the same person. While the average menstrual cycle length is considered to be around 28 days, it is important to note that normal cycle lengths can range anywhere from 21 to 35 days. Monitoring and tracking the menstrual cycle length can provide insights into a woman’s reproductive health, fertility, and hormonal patterns. It helps individuals understand their menstrual patterns, identify potential irregularities or changes, and plan accordingly for future menstrual periods and family planning.

The ovulation day is a key factor in predicting the menstrual cycle length of a woman. Ovulation refers to the release of an egg from the ovary, marking the most fertile phase of a woman’s menstrual cycle. The ovulation day represents the specific day on which ovulation is expected to occur.

**Exploratory Data Analysis (EDA)**:

Here are the activities of the EDA of this research.

Jupyter note 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

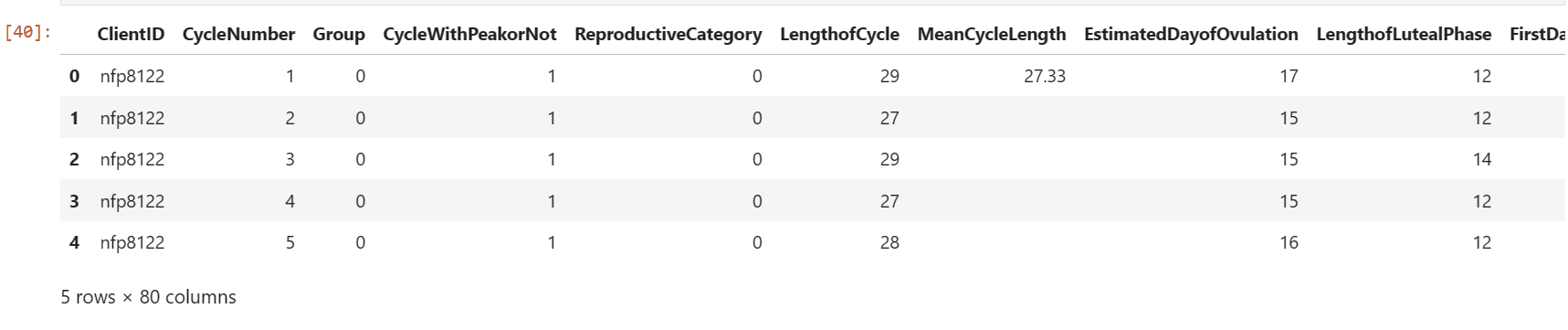
The dataset is loaded using this;

#Import the dataset

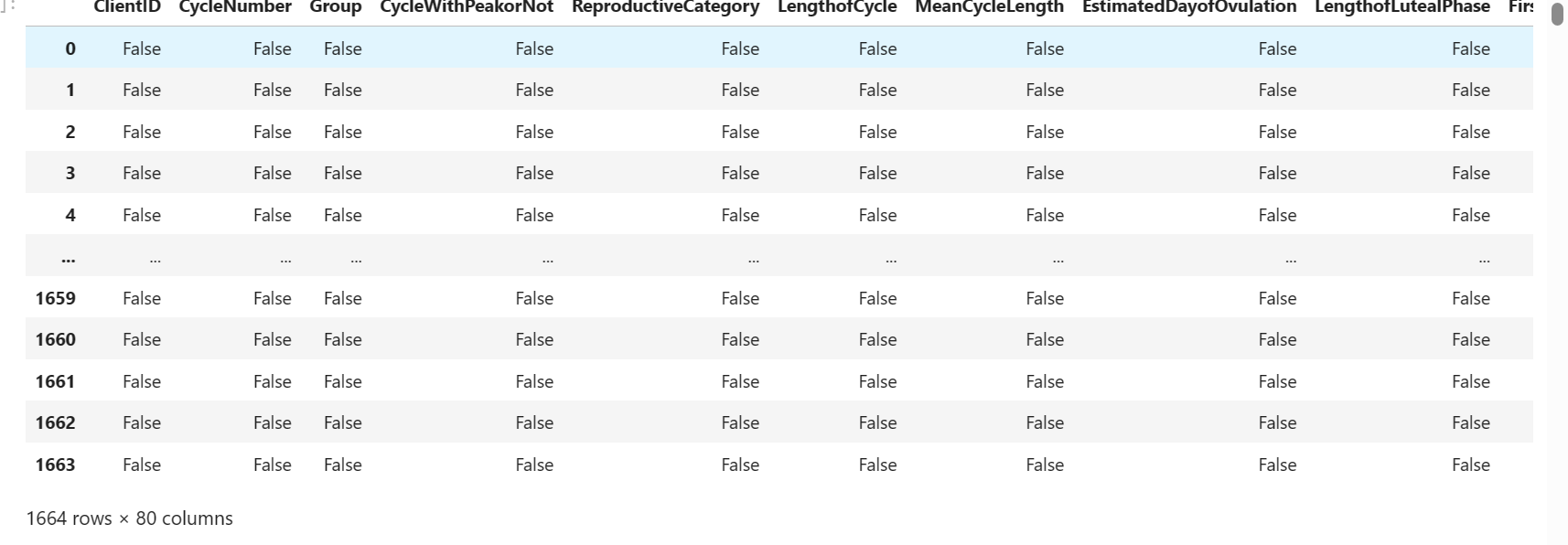
df = pd.read\_csv("C:/Users/hp/Desktop/MLs/DeployModelOvuLength/FedCycleData.csv")

#To view five rows out of the dataset

df.head()

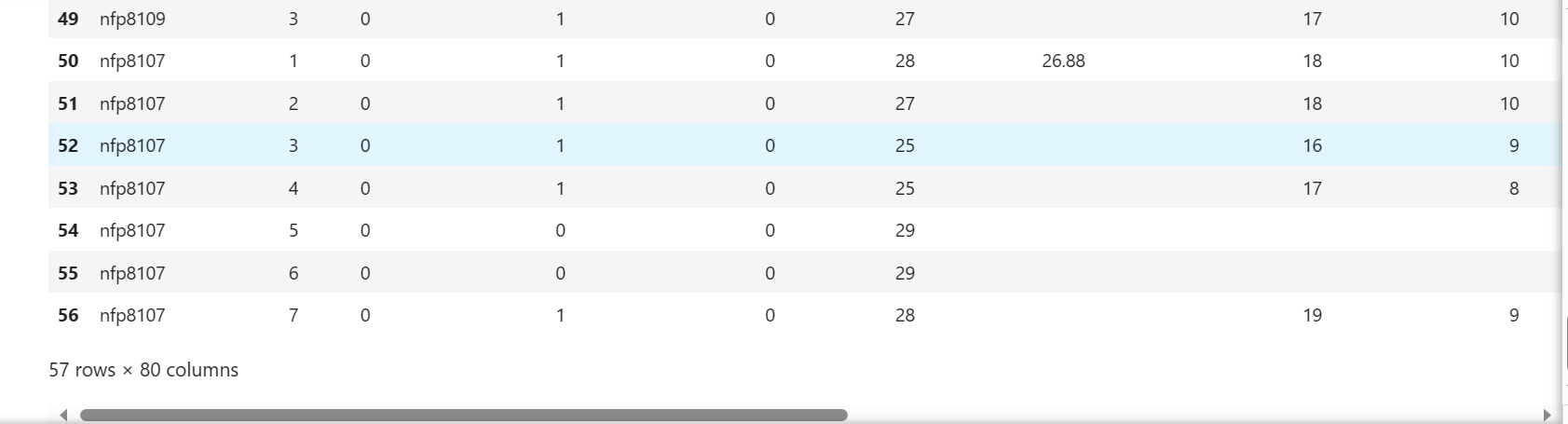


df.isnull()



# 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 were chosen** as appropriate machine learning algorithms and models for your 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, the three machine learning algorithm used include: Decision Tree, and Linear Regression Algorithm, Random Forest,

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 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 menstrual cycle length. Three different machine learning models were built to perform the prediction: linear regression, Random Forest , and Decision Tree. Among the three models, Decision tree model has the highest R2 score of 0.58 and was used as the best model for the prediction. Also, the model was deployed on a web application using Django.

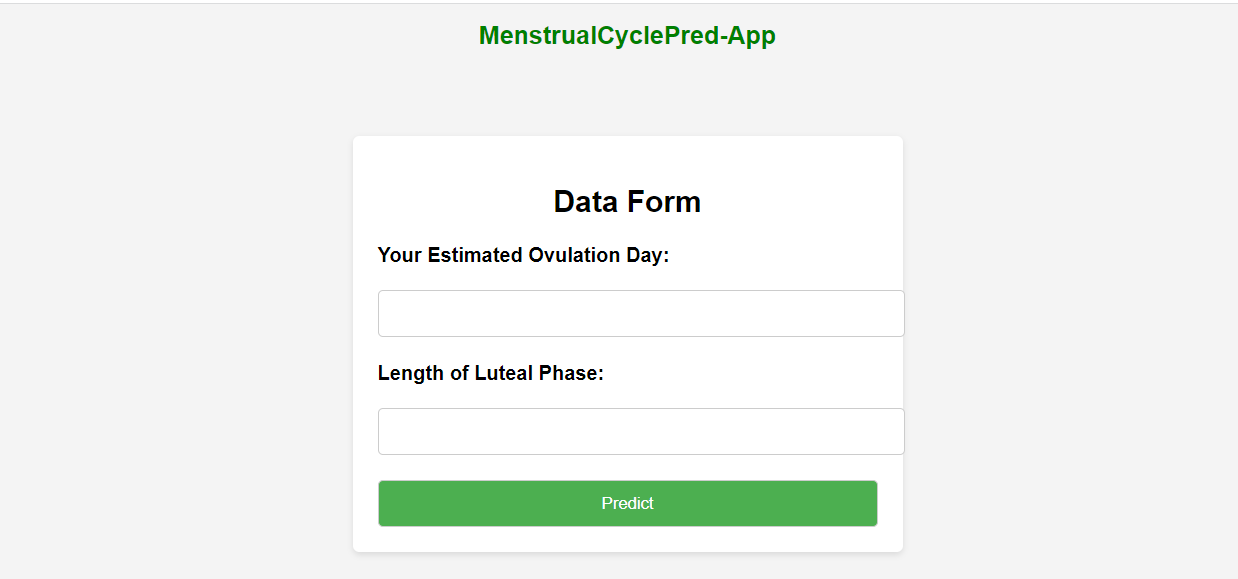
The introduction of the new system signifies a notable advancement over the existing framework, characterized by the introduction of a user-friendly interface that offers a seamless experience for users to predict their menstrual cycle length accurately. This innovation addresses a crucial need in women's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

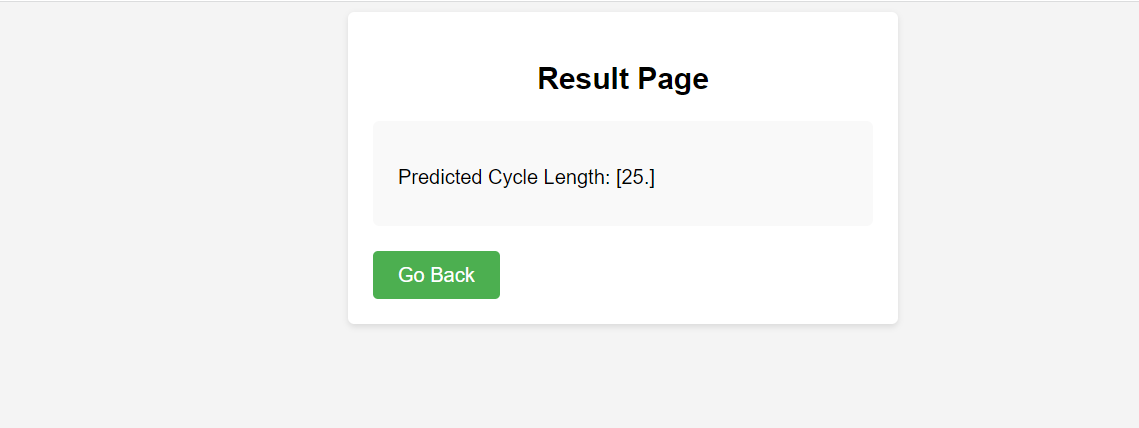
The development process entailed the creation of three distinct machine learning models, each specifically designed for accurate prediction: linear regression, Random Forest, and Decision Tree. These models, driven by advanced algorithmic methodologies, constitute the core of the prediction mechanism. 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 0.58. 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 0.58 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 seamless integration of the Decision Tree model into a web application environment adds an extra layer of accessibility and usability. This is accomplished through the utilization of the Django framework, a versatile and efficient web development tool. Deploying the model within a web application expands its reach and usability, allowing users to interact with it conveniently and obtain real-time predictions effortlessly. The interface provides 2 input fields denoting the Ovulation Day and length of luteal phase respectively. The user provides information regarding the estimated Ovulation Day and Luteal phase and clicks the predict button to reveal the predicted cycle length.





**Results**

Table 1: Proposed System Results Using only Three Features (feature[LengthofLutealPhase, EstimatedDayofOvulation]target[LengthofCycle]).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SN** | **MODEL** | **MAE** | **MSE** | **RMSE** | **R-squared (R²) Score** |
| 1 | Decision Tree Regressor | 0.52436 | 2.64024 | 1.62488 | 0.81327 |
| 2 | Random Forest | 0.528 | 3.008 | 1.73436 | 0.78726 |
| 3 | Linear Regression | 0.64695 | 2.29411 | 1.51463 | 0.83775 |
| 4 | Huber Regression | 0.364000 | 2.32399 | 1.52446 | 0.83564 |
| 5 | Least Angle Regression | 0.64695 | 2.29410 | 1.51463 | 0.83775 |
| 6 | Lasso Regresion | 0.934541 | 2.63912 | 1.51463 | 0.81335 |
| 7 | Lasso Least Angle Regression | 0.93453 | 2.63911 | 1.62453 | 0.81335 |
| 8 | **\***Dummy Regressor | 2.817227 | 14.13958 | 3.76026 | -2.88523 |
| 9 | Ridge Regressor | 0.64705 | 2.29409 | 1.51462 | 0.83775 |
| 10 | K-Nearest Neighbour | 0.602666 | 3.21076 | 1.79186 | 0.77292 |
| 11 | Elastic Net | 0.89844 | 2.56373 | 1.60116 | 0.81868 |
| 12 | Orthogonal Matching Pursuit | 0.64695 | 2.29410 | 1.51463 | 0.83775 |
| 13 | LightGradient Boosting Matching | 0.65289 | 3.92754 | 1.98180 | 0.72223 |
| 14 | Bayesian Ridge | 0.64738 | 2.29404 | 1.51461 | 0.83775 |
| 15 | Passive Aggressive Regression | 1.47201 | 4.80156 | 2.19124 | 0.82932 |
| 16 | Catboost | 1.47201 | 2.56321 | 1.60100 | 0.81872 |
| 17 | Gradient boosting | 0.577108 | 2.546139 | 1.595668 | 0.81992 |
| 18 | Extra Tree Regresor | 1.50789 | 4.51814 | 2.12559 | 0.68046 |
| 19 | Xgb | 0.59595 | 2.56644 | 1.6020136 | 0.81849 |

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 |
| 5 | Best fit Machine Learning Model. | Decision Tree Regressor |
| 7 | R2 value of the best fit Machine Learning Model | 0.58 |

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 | Linear Regression and Support Vector Machine | Logistic Regression, Decision tree Regression, and Random Forest Regression |
| 5 | R2 Score | 0.37 | 0.58 |
| 6 | Best Machine Learning Model | Linear Regression | Decision Tree regression |

Discussion

This section presents the results and discuss the findings of the research project focused on developing a new system for menstrual cycle length prediction. The system utilizes AI-powered algorithms, including Decision Tree, Random Forest, and Logistic Regression, to accurately forecast the menstrual cycle length of women. Figure 3.2 displays the architectural design of the new system. It showcases the components and their interactions, highlighting the required data (Ovulation Day, and length of luteal phase) that are used for predicting the cycle length. Figure 3.3 represents the result page of the new system when a user inputs the ovulation day and length of the luteal phase. This user-friendly interface allows women to receive prompt predictions of their menstrual cycle length, empowering them with valuable information for family planning and reproductive health management. Figure 3.4 represents the flowchart of the new system, depicting the sequential steps involved in predicting the menstrual cycle length. It highlights the data inputs, algorithmic processes, and final output generation. Table 1 showcases accuracy parameters for the three models, namely Decision Tree, Random Forest, and Logistic Regression. 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. Figure 4.1, Figure 4.2 and Figure 4.3 depict the comparison of actual values of the training dataset versus predicted values of the dataset for each model. These visual representations demonstrate how well the models approximate the real menstrual cycle lengths. The closer the predicted values align with the actual values, the more accurate the predictions are. From the three graphs, the predicted values of the Decision Tree model is closer to the actual values. This is due to it having the highest R2 score. Figure 4.4 displays a comparison of the R2 scores among the three models. The R2 score measures the proportion of the variance in the menstrual cycle length 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 menstrual cycle length 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 menstrual cycle length prediction app with an R2 score of 0.58 lies in its ability to provide a predictive model that can estimate the length of a woman’s menstrual cycle. This contributes to women’s reproductive health by offering insights into their cycle patterns and helping them plan their daily activities, anticipate the timing of their periods, and track their fertility. The R2 score of 0.58 indicates that the developed prediction model explains 58% of the variance in the menstrual cycle length based on the input variables “Length of luteal phase” and “Estimated ovulation day”. This implies that the model has moderate predictive power, capturing a significant portion of the variability in cycle length based on the available data. By deploying this menstrual cycle length 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 (such as estimated ovulation day and length of the luteal phase) and obtaining cycle length 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**

1. In conclusion, the researchers have successfully developed and deployed a web application using Django for an AI-powered fertility prediction system. The aim of the research was to accurately forecast the menstrual cycle length of women using three algorithms: Decision Tree regression, Random Forest Regression, and Logistic Regression. Through rigorous analysis and evaluation, it was determined that the Decision Tree regressor algorithm outperformed the others, achieving the highest R2 score. By deploying this system on a web application, we have made the fertility prediction capabilities easily accessible to women. The application provides a user-friendly interface where women can input relevant data, such as ovulation day and length of the luteal phase, to obtain accurate predictions of their menstrual cycle length. The successful deployment of the system on a web application using Django opens up opportunities for further enhancements and future developments. It can be expanded to include additional features such as personalized insights, fertility tracking and reminders. User feedback and engagement will be crucial for continuous improvement and refinement of the system. The AI-powered fertility prediction system, deployed as a web application using Django, represents a significant advancement in reproductive health technology. It provides women with a convenient and reliable tool to predict their menstrual cycle length, empowering them to take control of their fertility and make informed decisions about their reproductive well-being. The application will enable women to track the length of their menstrual cycles over time. With this data, they can gain insight into their menstrual patterns, identify any irregularities or abnormalities, and also plan for future menstrual cycles. Knowing their menstrual cycle length is essential for fertility awareness and family planning. In furtherance, the menstrual cycle length can be a valuable tool for monitoring overall reproductive health. With this, the users can know that consistent and regular cycles indicate a healthy hormonal balance, while irregularities in cycle length may indicate underlying health issues that require attention.

**Reference**

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.

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

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

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.

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

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.