A Project Report

on

MENSTRUAL CYCLE-FLOW CALCULATOR

Submitted in partial fulfillment of the

requirement for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

Under The Supervision of
Mrs. IndraKumari
Associate Assistant Professor
Department of Computer Science and Engineering

Submitted By

19SCSE1010598 - Ritika Jaiswal
19SCSE1010134 - Harshit Kumar

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
DECEMBER - 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the project, entitled **"MENSTRUAL CYCLE-FLOW CALCULATOR"** in partial fulfilment of the requirements for the award of the BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of **JULY-2021 to DECEMBER-2021**, under the supervision of **MS. INDRAKUMARI, ASSISTANT PROFESSOR Department of Computer Science and Engineering**, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

19SCSE10100598 RITIKA JAISWAL

19SCSE1010134 HARSHIT KUMAR

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor

(MS. INDRAKUMARI ASSISTANT PROFESSOR)

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of **19SCSE1010134 HARSHIT KUMAR, 19SCSE1010598 RITIKA JAISWAL** has been held on _____ and his/ her work is recommended for the award of **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING.**

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date:

Place: Greater Noida

Abstract

CYCLEFlow is a machine learning based application that helps women to get to know the needs of their body and hygiene by Predicting their Menstrual Cycle Ovulation day , and many more things . So that they can have a better , more conscious and Comfortable Life.

Health : higher chance to early stage detection and Health focused life

Family planning : Protection and fertility .

Daily life : easier planning the daily routine.

Features : Current app can predict Day of Ovulation by taking some detail such as length of menstrual cycle etc by inputs from user.

Upcoming Features:

1.Unique menstrual cycle prediction for women to check the big parts of next period such as beginning and ending of menstrual cycle, beginning and ending fertile or non-fertile days

2.Warning for potential health problems based on the irregularity of menstrual cycle.

Table of Content

CHAPTER 1: Introduction

CYCLEFlow is a machine learning based application that helps women to get to know the needs of their body and hygiene by Predicting their Menstrual Cycle Ovulation day , and many more things. So that they can have a better, more conscious and Comfortable Life. As we know, the menstrual cycle is one of the most dynamic and unpredictable health conditions that women tackle in their day to day lives.

Forecast of the length of the feminine cycle and of its phases the pre ovular or follicular stage, the ovulation, and the postovular or luteal phase is a key stage in fruitlessness the executives (Stanford and others, 2003) and normal family arranging (see, for example Knight and Clubb, 1996). Distinguishing proof of a lady's prolific window is generally founded on schedule computations (Ogino, 1930; Colombo and Scarpa, 1996; Lamprecht and Grummer-Strawn, 1996; Arevalo and others, 1999; Arevalo and others, 2002), now and then joined with other self-checking side effects, like basal internal heat level or vulvar perception of cervical bodily fluid. Also, the exactness of certain strategies, for instance, postcoital tests, just as the accomplishment of certain treatments are reliant upon satisfactory planning of ovulation. Understanding the time advancement of the length of the whole cycle and of its stages is a significant issue likewise in regenerative and general wellbeing examines, as feminine brokenness apparently is identified with diminished fruitfulness and expanded future danger of different constant infections, for example, bosom malignant growth, cardiovascular illnesses, and diabetes (Yen, 1991; Harlow and Ephross, 1995; Guo and others, 2006).

Utilizing rehashed estimations of the length of the whole cycle and of the follicular stage given by a huge English data set, our point is to foster a measurable model that empowers forecast of every one of these 2 period qualities for a lady. Numerous factual models have been proposed in the writing for cycle lengths. Harlow and Matanoski (1991) and Harlow and Zeger (1991) characterize cycles into 2 gatherings, standard and nonstandard, characterized as having lengths underneath or over 44 days, individually, and look at covariate consequences for the mean length of standard cycles through a direct blended model. Lin and others (1997) stretch out this model to represent the heterogeneity between ladies, while Harlow and others (2000) assess the age impact on the likelihood of having a nonstandard cycle utilizing a summed up assessing

condition. Guo and others (2006) propose a combination of a typical dissemination and a moved Weibull circulation to show both customary and bizarre cycle lengths.

Every one of the past references, and the majority of the accessible writing, center around the minimal appropriation of cycle length. To consider unequivocally the longitudinal idea of the information, we propose a powerful methodology for both the whole cycle length and the hour of ovulation. We utilize a state-space process (West and Harrison, 1997) to depict the fleeting conduct of the series of lengths for every lady. Our plan is driven by the highlights of the peculiarity that are known from past investigations or got from organic contemplations and examination of the information. To catch the inconstancy across ladies, the singular cycles are implanted into a multivariate framework through a Bayesian progression wherein model boundaries are permitted to change across subjects as per a predefined likelihood dissemination. The subsequent Bayesian various leveled dynamic model enjoys the benefit of empowering an exchange of data across subjects, making up for the moderately short history of data accessible on every lady and the assessment of populace boundaries that can be utilized for forecasts on ladies excluded from the data set. Moreover, however the actual model is non-Gaussian and nonlinear, restrictively on a reasonably increased boundary space, it turns into a progressive straight Gaussian state-space model on which derivation can be made by means of the Gibbs methods of Shephard (1994) and Carter and Kohn (1994). Consequently, back and prescient estimations can be done effectively and utilized in applications. For instance, we will show that the draws from the prescient conveyance of the pre ovular stage length can be utilized to gauge the likelihood of origination in the following cycle as a component of coital conduct and that this chance prompts intriguing applications for the ID of the most rich window of a lady.

The design of the paper is as per the following. Area 2 gives an itemized depiction of the informational collection. In Section 3, the Bayesian progressive powerful model is created. In Section 4, a Markov chain Monte Carlo (MCMC) calculation for back estimations is examined. In Section 5, the model is fitted to the English information base and indicative looks at are conveyed. Segment 6 shows how the fitted model can be utilized to gauge the likelihood of origination in later cycles as an element of the intercourse design. Segment 7 contains a conversation of the outcomes and a few thoughts for future work.

CHAPTER 2: Literature Review

2.1 The paper intends to:

This study aimed to describe differences in menstrual cycle length, variability, and menstrual phase across women of different ages . We also reported on demographic and lifestyle characteristics across the median cycle length. The study analyses will provide extensive worldwide evidence on the characteristics of menstrual cycle length and patterns among a global cohort of Flo app users. This information is necessary to support recommendations within current obstetric clinical guidelines around menstrual cycle length and patterns for clinical use in fertility programs.

2.2 Methodology:

The project is concerned about building a mensural cycle and ovulation day prediction model using the machine learning algorithms. The project is constant developing in different phase because the focus of it towards model development in a machine learning interface using Jupitor notebooks.
Machine learning generally requires a fair amount of time in training the model and testing it through different parameters and also a good quality of training dataset. In other words if we're saying the model is considered pretty much as good in phase 1 , if the model perform in phase 2.

2.3 Menstrual Cycle Characteristics

Women manually logged information about their menstrual cycles (days of menstruation), including the intensity of menstrual flow (light, medium, heavy). The start and end dates of menstrual cycles were defined by the logged first day of menstruation. To estimate mean and median cycle lengths, we computed women inputs and then calculated the mean and median for population groups.

2.4 Model Assessment

Current app can predict Day of Ovulation by taking some detail such as length of menstrual cycle etc by inputs from user .

Upcoming Features:

1. unique menstrual cycle prediction for women to check the big parts of next period such as beginning and ending of menstrual cycle, beginning and ending fertile or non-fertile days
2. warning for potential health problems based on the irregularity of menstrual cycle.

2.5 Proposed Model Interface :

Tools and Technology Used

Various python libraries:

numpy

pandas

itertools

matplotlib

sklearn

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modelling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy,
SciPy and Matplotlib.

Prerequisites

Before we start using scikit-learn latest release, we require the following −

- Python (>=3.5)
- NumPy (>= 1.11.0)
- Scipy (>= 0.17.0)li
- Joblib (>= 0.11)
- Matplotlib (>= 1.5.1) is required for Sklearn plotting capabilities.

- Pandas (>= 0.18.0) is required for some of the scikit-learn examples using data structure and analysis.

2.6 Existing Work

Model 1:Sequential predictions of menstrual cycle lengths

Consider an individual woman, and let yt denote the length in days of her tth menstrual cycle (t=1,2,…) or, alternatively, of the hypothermic phase in the basal body temperature. Our main objective is the derivation of the one-step ahead predictive distribution
graphic

$$F_{t+1}(x) = P\{y_{t+1} \leqslant x | y_1, \ldots, y_t\},$$

for t=1,2,…. We are interested in evaluating Ft+1(x) in its entirety since the nonnegligible variability within individual series means that interval forecasts are more appropriate than point predictions (Marshall, 1965).
In looking for a statistical model suitable to describe the observed processes, we have to take into account that there is variability both within and between-women. The approach used to model these 2 sources of variation can be described as follows. To explain variations from cycle to cycle for a particular woman, we propose a parametric model, that is,

$$P\{y_{t+1} \leqslant x | y_1, \ldots, y_t, \theta\} = G_{t+1}(x | y_1, \ldots, y_t, \theta),$$

where Gt+1 is fully specified and θ is a vector of unknown parameters. We then assume that the functional form of Gt+1 is the same for all women, while the residual variability between women can be described by allowing θ to vary across subjects according to a probability distribution p(θ| ζ). Finally, we cast the problem in a Bayesian framework by specifying a prior distribution for ζ.

2.7 About:

Every month during the years between pubescence and menopause, a woman's body goes through various changes to prepare her for a potential pregnancy. This series of chemical-driven occasions is known as the period.

During each period, an egg creates and is let out of the ovaries. The arranging of the uterus fabricates. In the event that pregnancy doesn't occur, the uterine covering sheds during a feminine period. Then, at that point, the cycle begins once more.

A woman's period is isolated into four stages:

- feminine stage
- follicular stage
- ovulation stage
- luteal stage

The length of each stage can vary from one lady to another, and it can change after some time.

2.8 Feminine stage

The feminine stage is the primary phase of the period. It's likewise when you get your period.

This stage begins when an egg from the past cycle isn't prepared. Since pregnancy hasn't occurred, levels of the chemicals estrogen and progesterone drop.

The thickened covering of your uterus, which would uphold a pregnancy, is not generally required, so it sheds through your vagina. During your period, you discharge a blend of blood, bodily fluid, and tissue from your uterus.

You might have period indications like these:

cramps (attempt these home cures)

delicate bosoms

swelling

state of mind swings

touchiness

cerebral pains

sleepiness

low back torment

Overall, ladies are in the feminine period of their cycle for 3 to 7 days. A few ladies have longer periods than others.

2.9 Follicular stage

The follicular stage begins the principal day of your period (so there is some cross-over with the feminine stage) and closures when you ovulate.

It begins when the nerve center conveys a message to your pituitary organ to deliver a follicle-animating chemical (FSH). This chemical animates your ovaries to deliver around 5 to 20 little sacs called follicles. Every follicle contains a juvenile egg.

Hands down the best egg will ultimately develop. (On uncommon events, a lady might have two eggs mature.) The remainder of the follicles will be reabsorbed into your body.

The developing follicle sets off a flood in estrogen that thickens the coating of your uterus. This establishes a supplement-rich climate for an undeveloped organism to develop.

The normal follicular phrase trusted Source goes on for around 16 days. It can go from 11 to 27 days, contingent upon your cycle.

2.10 Ovulation stage

Rising estrogen levels during the follicular stage trigger your pituitary organ to deliver luteinizing chemicals (LH). This is which begins the course of ovulation.

Ovulation is the point at which your ovary delivers a developed egg. The egg goes down the fallopian tube toward the uterus to be treated by sperm.

The ovulation stage is the possible time during your period when you can get pregnant. You can tell that you're ovulating by side effects like these:

a slight ascent in basal internal heat level

thicker release that has the surface of egg whites

Ovulation occurs at around day 14 if you have a 28-day cycle solidly in the center of your period. It goes on around 24 hours. Following a day, the egg will pass on or disintegrate on the off chance that it isn't prepared.

DID YOU KNOW?
Since sperm can satisfy five days, pregnancy can happen if a lady engages in sexual relations as much as five days preceding ovulation.

2.11 Luteal stage

After the follicle delivers its egg, it changes into the corpus luteum. This design discharges chemicals, essentially progesterone and some estrogen. The ascent in chemicals saves your uterine covering thick and prepared for a treated egg to embed.

In the event that you do get pregnant, your body will deliver human chorionic gonadotropin (hCG). This is the chemical pregnancy tests recognize. It keeps up with the corpus luteum and keeps the uterine covering thick.

If you don't get pregnant, the corpus luteum will contract away and be resorbed. This prompts diminished degrees of estrogen and progesterone, which causes the beginning of your period. The uterine covering will be shed during your period.

During this stage, on the off chance that you don't get pregnant, you might encounter manifestations of premenstrual disorder (PMS). These include:

- bulging
- bosom enlarging, agony, or delicacy
- mind-set changes
- migraine
- weight gain
- changes in sexual craving
- food yearnings

- inconvenience resting

The luteal stage goes on for 11 to 17 days. The normal lengthTrusted Source is 14 days.

2.12 Distinguishing normal issues

Each woman's period is unique. A few ladies get their period simultaneously every month. Others are more unpredictable. A few ladies drain all the more intensely or for a more extended number of days than others.

Your feminine cycle can likewise change during specific occasions of your life. For instance, it can get more unpredictable as you draw near to menopause.

One method for seeing whether you're having any issues with your monthly cycle is to follow your periods. Record when they start and end. Additionally record any progressions to the sum or number of days you drain, and regardless of whether you have spotting between periods.

2.13 Any of these things can adjust your period:

Anti-conception medication. The anti-conception medication pill might make your periods more limited and lighter. While on certain pills, you wont get a period by any means.

Pregnancy. Your periods should quit during pregnancy. Missed periods are one of the most clear initially signs that youre pregnant.

Polycystic ovary condition (PCOS). This hormonal unevenness keeps an egg from growing ordinarily in the ovaries. PCOS causes unpredictable monthly cycles and missed periods.

Uterine fibroids. These noncancerous developments in your uterus can make your periods longer and heavier than expected.

Dietary problems. Anorexia, bulimia, and other dietary problems can upset your feminine cycle and make your periods stop.

The following are a couple of indications of an issue with your period:

You've skipped periods, or your periods have halted totally.

Your periods are sporadic.

You drain for over seven days.

Your periods are under 21 days or over 35 days separated.

You drain between periods (heavier than spotting).

On the off chance that you have these or different issues with your monthly cycle or periods, converse with your medical services supplier.

2.14 The action item

Each woman's period is unique. What's typical for you probably won't be typical for another person.

It's critical to get to know your cycle including when you get your periods and how long they last. Be ready for any changes, and report them to your medical care supplier.

CHAPTER 3 : METHODS

Hereby we describe the detailed description of the used datasets, algorithms and performance metrics for the final analysis and conclusion.

3.1 Supervised Machine Learning

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.

In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

In the real-world, supervised learning can be used for Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.

3.1.1 Steps Involved in Supervised Learning:

- First Determine the type of training dataset
- Collect/Gather the labelled training data.
- Split the training dataset into training dataset, test dataset, and validation dataset.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision

tree, etc.

- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

3.1.2 Types of supervised Machine learning Algorithms:

Supervised learning can be further divided into two types of problems:

(i) Regression

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning:

- Linear Regression
- Regression Trees
- Non-Linear Regression
- Bayesian Linear Regression
- Polynomial Regression

(ii) Classification

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

- Spam Filtering,
- Random Forest
- Decision Trees
- Logistic Regression
- Support vector Machines

3.2 Unsupervised Machine Learning

In the previous topic, we learned supervised machine learning in which models are trained using labeled data under the supervision of training data. But there may be many cases in which we do not have labeled data and need to find the hidden patterns from the given dataset. So, to solve such types of cases in machine learning, we need unsupervised learning techniques.

3.2.1 What is Unsupervised Learning?

As the name suggests, unsupervised learning is a machine learning technique in which models are not supervised using training dataset. Instead, models itself find the hidden patterns and insights from the given data. It can be compared to learning which takes place in the human brain while learning new things. It can be defined as:

Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision.

Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data. The goal of unsupervised learning is to find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.

3.2.2 Why use Unsupervised Learning?

Below are some main reasons which describe the importance of Unsupervised Learning:

- Unsupervised learning is helpful for finding useful insights from the data.
- Unsupervised learning is much similar as a human learns to think by their own experiences, which makes it closer to the real AI.
- Unsupervised learning works on unlabeled and uncategorized data which make unsupervised learning more important.
- In real-world, we do not always have input data with the corresponding output so to solve such cases, we need unsupervised learning.

3.3Regression Analysis in Machine learning

Regression analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables. More specifically, Regression analysis helps us to understand how the value of the dependent variable is changing corresponding to an independent variable when other independent variables are held fixed. It predicts continuous/real values such as temperature, age, salary, price, etc.

Regression is a supervised learning technique which helps in finding the correlation between variables and enables us to predict the continuous output variable based on the one or more predictor variables. It is mainly used for prediction, forecasting, time series modeling, and determining the causal-effect relationship between variables.

In Regression, we plot a graph between the variables which best fits the given datapoints, using this plot, the machine learning model can make predictions about the data. In simple words, "Regression shows a line or curve that passes through all the datapoints on target-predictor graph

in such a way that the vertical distance between the datapoints and the regression line is minimum." The distance between datapoints and line tells whether a model has captured a strong relationship or not.

Some examples of regression can be as:

- Prediction of rain using temperature and other factors
- Determining Market trends
- Prediction of road accidents due to rash driving.

3.3.1 Terminologies Related to the Regression Analysis:

- **Dependent Variable:** The main factor in Regression analysis which we want to predict or understand is called the dependent variable. It is also called target variable.
- **Independent Variable:** The factors which affect the dependent variables or which are used to predict the values of the dependent variables are called independent variable, also called as a predictor.
- **Outliers:** Outlier is an observation which contains either very low value or very high value in comparison to other observed values. An outlier may hamper the result, so it should be avoided.
- **Multicollinearity:** If the independent variables are highly correlated with each other than other variables, then such condition is called Multicollinearity. It should not be present in the dataset, because it creates problem while ranking the most affecting variable.
- **Underfitting and Overfitting:** If our algorithm works well with the training dataset but not well with test dataset, then such problem is called Overfitting. And if our algorithm does not perform well even with training dataset, then such problem is called underfitting.

3.3.2 Why do we use Regression Analysis?

As mentioned above, Regression analysis helps in the prediction of a continuous variable. There are various scenarios in the real world where we need some future predictions such as weather condition, sales prediction, marketing trends, etc., for such case we need some technology which can make predictions more accurately. So for such case we need Regression analysis which is a statistical method and used in machine learning and data science. Below are some other reasons for using Regression analysis:

- Regression estimates the relationship between the target and the independent variable.
- It is used to find the trends in data.
- It helps to predict real/continuous values.
- By performing the regression, we can confidently determine the most important factor, the least important factor, and how each factor is affecting the other factors.

3.3.3 Types of Regression

There are various types of regressions which are used in data science and machine learning. Each type has its own importance on different scenarios, but at the core, all the regression methods analyze the effect of the independent variable on dependent variables. Here we are discussing some important types of regression which are given below:

- Linear Regression
- Logistic Regression
- Polynomial Regression
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression

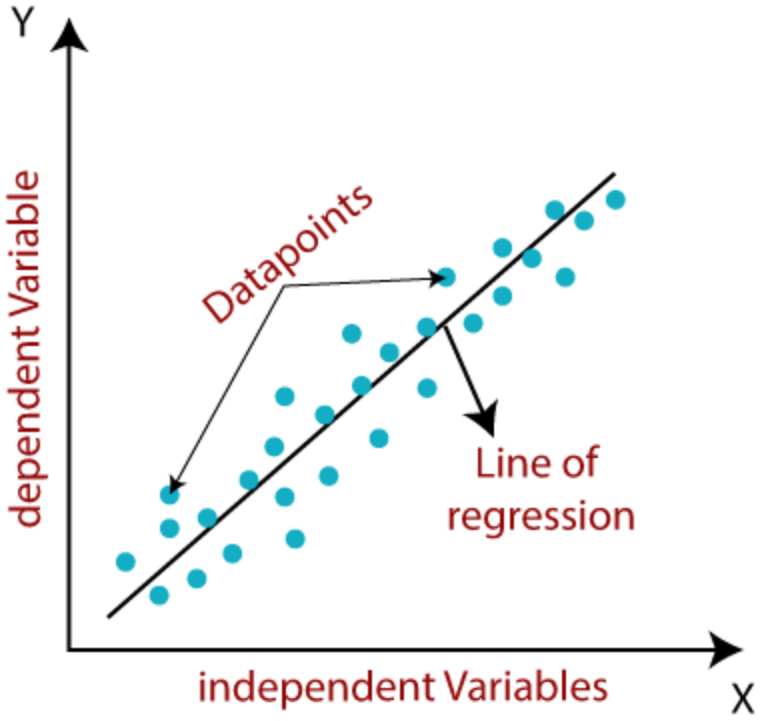
- Ridge Regression
- Lasso Regression:

3.4 Linear Regression in Machine Learning

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:



Mathematically, we can represent a linear regression as:

$y= a0+a1x+ \varepsilon$

Here,

Y= Dependent Variable (Target Variable)

X= Independent Variable (predictor Variable)

a0= intercept of the line (Gives an additional degree of freedom)

a1 = Linear regression coefficient (scale factor to each input value).

ε = random error

The values for x and y variables are training datasets for Linear Regression model

representation.

3.4.1 Types of Linear Regression

Linear regression can be further divided into two types of the algorithm:

- *Simple Linear Regression:*

  If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

- *Multiple Linear regression:*

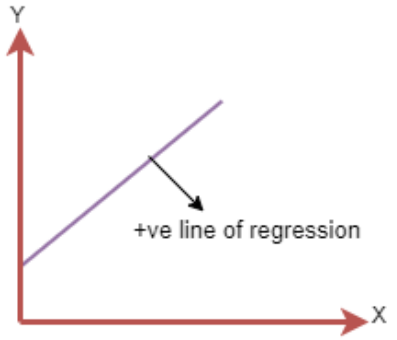
  If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

3.4.2 Linear Regression Line

A linear line showing the relationship between the dependent and independent variables is called a regression line. A regression line can show two types of relationship:

- *Positive Linear Relationship:*

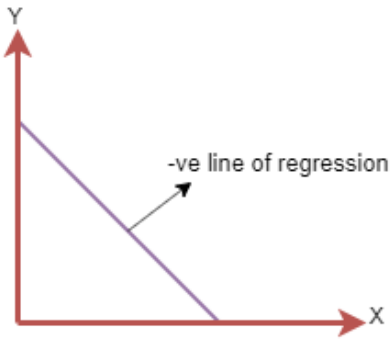
  If the dependent variable increases on the Y-axis and independent variable increases on X-axis, then such a relationship is termed as a Positive linear relationship.



The line equation will be: $Y= a_0+a_1x$

- *Negative Linear Relationship:*

  If the dependent variable decreases on the Y-axis and independent variable increases on the X-axis, then such a relationship is called a negative linear relationship.



The line of equation will be: $Y= -a_0+a_1x$

3.4.3 Finding the best fit line:

When working with linear regression, our main goal is to find the best fit line that means the error between predicted values and actual values should be minimized. The best fit line will have the least error.

The different values for weights or the coefficient of lines (a0, a1) gives a different line of

regression, so we need to calculate the best values for a0 and a1 to find the best fit line, so to calculate this we use cost function.

3.4.4 Cost function-

- The different values for weights or coefficient of lines (a0, a1) gives the different line of regression, and the cost function is used to estimate the values of the coefficient for the best fit line.
- Cost function optimizes the regression coefficients or weights. It measures how a linear regression model is performing.
- We can use the cost function to find the accuracy of the mapping function, which maps the input variable to the output variable. This mapping function is also known as Hypothesis function.

For Linear Regression, we use the Mean Squared Error (MSE) cost function, which is the average of squared error occurred between the predicted values and actual values. It can be written as:

For the above linear equation, MSE can be calculated as:

$$MSE = 1\frac{1}{N}\sum_{i=1}^{n}(y_i - (a_1 x_i + a_0))^2$$

Where,

N=Total number of observation

Yi = Actual value

(a1xi+a0)= Predicted value.

Residuals: The distance between the actual value and predicted values is called residual. If the observed points are far from the regression line, then the residual will be high, and so cost function will high. If the scatter points are close to the regression line, then the residual will be small and hence the cost function.

3.4.5 Gradient Descent:

- Gradient descent is used to minimize the MSE by calculating the gradient of the cost function.
- A regression model uses gradient descent to update the coefficients of the line by reducing the cost function.
- It is done by a random selection of values of coefficient and then iteratively update the values to reach the minimum cost function.

3.4.6 Model Performance:

The Goodness of fit determines how the line of regression fits the set of observations. The process of finding the best model out of various models is called optimization. It can be achieved by below method:

(i) R-squared method:

- R-squared is a statistical method that determines the goodness of fit.
- It measures the strength of the relationship between the dependent and independent variables on a scale of 0-100%.
- The high value of R-square determines the less difference between the predicted values and actual values and hence represents a good model.

- It is also called a coefficient of determination, or coefficient of multiple determination for multiple regression.
- It can be calculated from the below formula:

$$\text{R-squared} = \frac{\text{Explained variation}}{\text{Total Variation}}$$

3.4.7 Assumptions of Linear Regression

Below are some important assumptions of Linear Regression. These are some formal checks while building a Linear Regression model, which ensures to get the best possible result from the given dataset.

- **Linear relationship between the features and target:**

  Linear regression assumes the linear relationship between the dependent and independent variables.

- **Small or no multicollinearity between the features:**

  Multicollinearity means high-correlation between the independent variables. Due to multicollinearity, it may difficult to find the true relationship between the predictors and target variables. Or we can say, it is difficult to determine which predictor variable is affecting the target variable and which is not. So, the model assumes either little or no multicollinearity between the features or independent variables.

- **Homoscedasticity Assumption:**

  Homoscedasticity is a situation when the error term is the same for all the values of independent variables. With homoscedasticity, there should be no clear pattern distribution of data in the scatter plot.

- **Normal distribution of error terms:**

  Linear regression assumes that the error term should follow the normal distribution pattern. If error terms are not normally distributed, then confidence intervals will become

either too wide or too narrow, which may cause difficulties in finding coefficients.

It can be checked using the q-q plot. If the plot shows a straight line without any deviation, which means the error is normally distributed.

- **No autocorrelations:**
  The linear regression model assumes no autocorrelation in error terms. If there will be any correlation in the error term, then it will drastically reduce the accuracy of the model. Autocorrelation usually occurs if there is a dependency between residual errors.

3.5 Heroku Development

Heroku is a container-based cloud Platform as a Service (PaaS). Developers use Heroku to deploy, manage, and scale modern apps. This platform is elegant, flexible, and easy to use, offering developers the simplest path to getting their apps to market.

Heroku is fully managed, giving developers the freedom to focus on their core product without the distraction of maintaining servers, hardware, or infrastructure. The Heroku experience provides services, tools, workflows, and polyglot support—all designed to enhance developer productivity.

Heroku is a cloud platform as a service (PaaS) supporting several programming languages. One of the first cloud platforms, Heroku has been in development since June 2007, when it supported only the Ruby programming language, but now supports Java, Node.js, Scala, Clojure, Python, PHP, and Go. For this reason, Heroku is said to be a polyglot platform as it has features for a developer to build, run and scale applications in a similar manner across most languages.

CHAPTER 4: THE DATA

Data on 1798 women have been collected from clients of the Catholic Marriage Advisory Council of England and Wales, whose centers provided free counseling and educational service

on fertility awareness and natural family planning. Women in the study, whose ages range between 18 and 50 years, were known to be healthy. Most women were also fertile, having already given birth to at least one child, and for those (about 20%) that had not yet conceived, there was no reason to doubt their fertility (Miolo *and others*, 1993). Each woman provided a sequence of at least 6 consecutive cycles, leading to a total of 36 641 cycles. The longest recorded sequence of consecutive cycles comprises 109 measurements. Some women contributed more than one sequence to the database, though we decided to consider only the first sequence available as we have no information on the reasons that caused this temporary dropout and fear that the inclusion of the following sequences might bias the analysis.

For each woman, the length of every menstrual cycle was recorded, together with the daily basal body temperature and the days during which menstrual bleeding occurred (Marshall, 1979). Following the World Health Organization standard, a menstrual cycle is defined as the interval from the first day of one bleeding episode up to and including the day before the next bleeding episode. Using the so-called "three over six rule" (Marshall, 1979), a conventional marker of the day of ovulation for each cycle was determined manually as the first time in the cycle that the minimum basal temperature over 3 consecutive days was above the maximum temperature over the 6 immediately preceding days (Miolo *and others*, 1993). The day of ovulation is taken as the last of the 6 days with lower temperature. However, for some cycles, the length of the pre ovular phase is missing either because the cycle is apparently monophasic or because the registration of the basal body temperature is incomplete. We decided to consider only sequences with at least 3 consecutive nonmissing values of the time of ovulation. Thus, the data used in the study of the follicular phase are a subset of the total data.

4.1 Data characteristics

Previous studies, biological considerations, and a rough inspection of the data set suggest that the individual sequences of cycle length and pre ovular phase length have some common features:

As data are recorded in days, observed lengths are discrete.

- A slow downward time trend is generally observed for sequences covering many years. It is well known (e.g. Treloar *and others*, 1967; Vollman, 1977) that the mean length tends

to decrease over a long period of time, being dependent on a woman's age. As an example, in Figure 1, the cycle length is plotted against time for the woman with the longest recorded sequence of consecutive cycles in the database.

- For some long sequences, there appears to be a sudden change in the mean level, which might be due to changes in the woman's life style or behavior.

- After accounting for a possible trend, a temporal dependence among observations seems to be present for many women. For the cycle length, there appears to be a negative lag-one autocorrelation which was previously noticed also by Colombo and Bassi (1996). This suggests that for a stationary sequence, a long cycle is most likely followed by a relatively short one and vice versa. As an example, for the same woman as in

- For some women, the data include very large or very small observations that can be regarded as outliers with respect to the woman's regular pattern. Some of these anomalous cycles might have a specific biological explanation, such as undetectable early losses.

- Heterogeneity is observed across women.

CHAPTER 5: CONCLUSION

5.1 Target Audience :

Women who want to minimise the chance of being pregnant, but do not have money for professional protection solution.

Women and their own partner who plan to have a baby.

Women who want to plan their daily routine based on predicted menstrual Cycle .

•Impact

Health : higher chance to early stage detection and Health focused life

Family planning : Protection and fertility .

Daily life : easier planning the daily routine.

biggest advantage of an application is that it does not need extra hardware. It can run on an existing server. Since the user-side is optional, if the woman has a mobile device (such as phone, tablet, laptop), she can use their own application.

5.2 Sustainability & Impact

This application offers a solution for an existing problem, as long as the problem exists our solution has a good reason to exist. Women can follow their own menstrual cycle and they can be warned about potential illnesses or any other abnormal factor of their cycle. Women can plan their life according to the menstrual cycle or to the fertile and non-fertile periods.