Team algoHERithm

Azure Women's Hackathon 2022
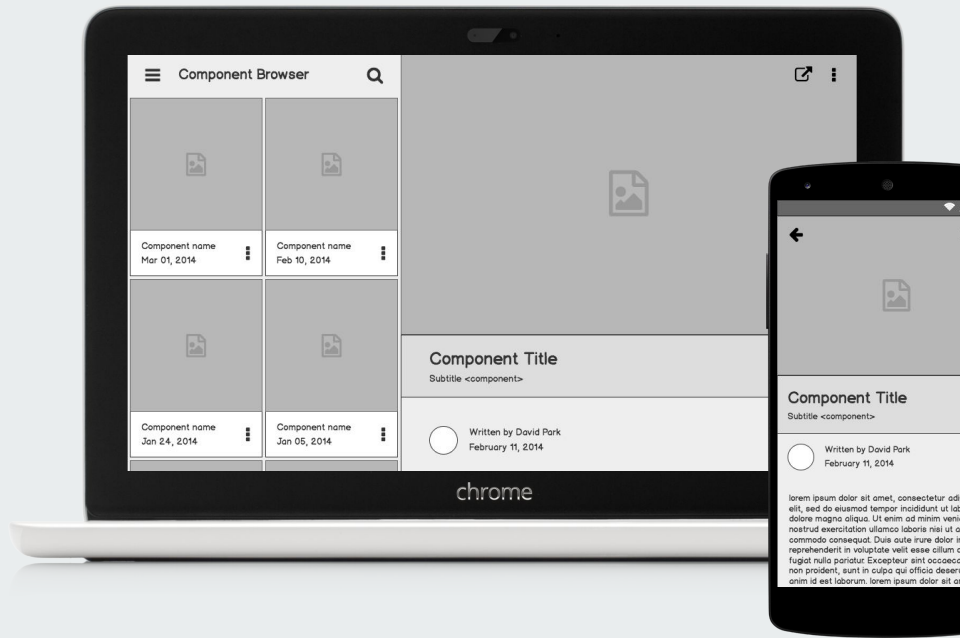
Outline

The Problem

Polycystic Ovary Syndrome or PCOS is an endocrine disorder that occurs in women of reproductive age. The condition once detected cannot be cured but treatment can help relieve its effects. The exact cause of PCOS is still unknown but there are certain factors that portray the risk of getting PCOS. The factors that result in this syndrome are: obesity, insulin resistance, blood pressure, depression, inflammation. While the symptoms include: hirsutism, Oligo-ovulation, acne, heavy bleeding, skin darkening. Using the causes and symptoms, a model is prepared in order to accept them as features and outputs the presence or absence of this condition

Problem Statement

Detecting PCOS using Machine Learning

Elaborative Problem

Women - Health

- Health: Without good health no achievements can be made
- An estimated one in five (20%) Indian women suffer from PCOS. If not monitored in time, the condition can have serious health impacts.

2021 Women's Health issue

Fibroids
2.4%

Difficulties conceiving
2.0%

Vaginal discharge
10.5%

Urinary Tract Infecti…
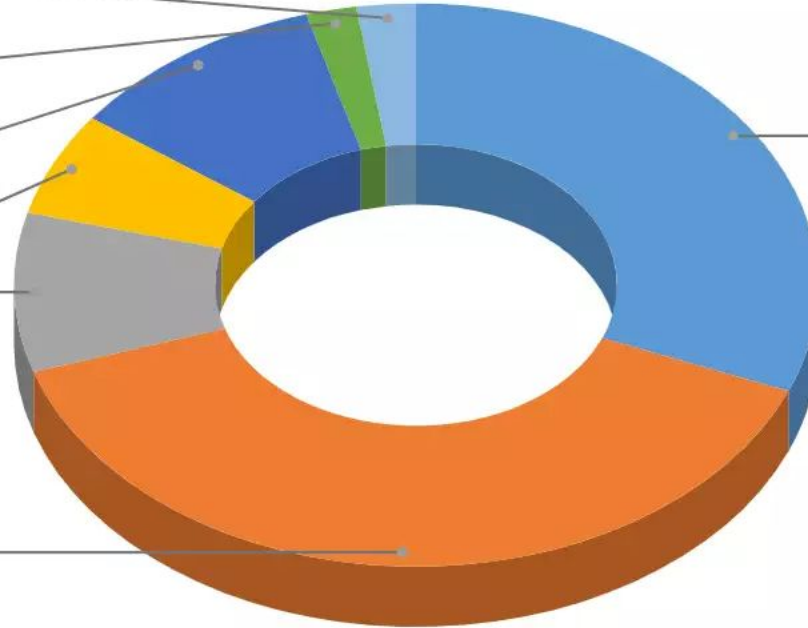6.0%

Breast pain
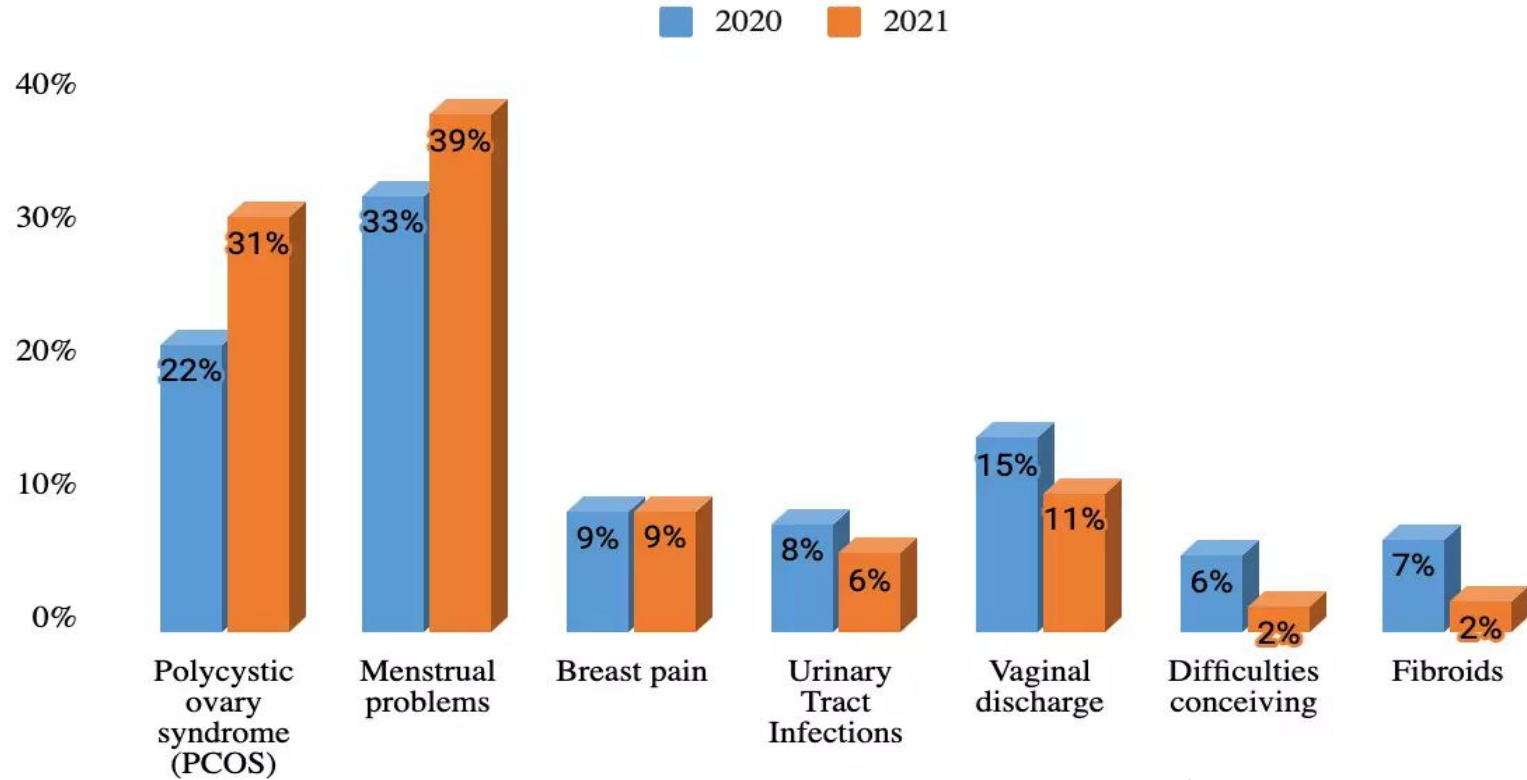9.2%

Menstrual problems
38.8%

Polycystic ovary synd…
31.2%

Women's Health Issue : A Comparison

Reference:
https://health.economictimes.indiatimes.com/

Solution Proposal

PCOS affects 4%–20% of women of reproductive age worldwide. The prevalence, diagnosis, etiology, management, clinical practices, psychological issues, and prevention are some of the most confusing aspects associated with PCOS. The aim is to exact relevant facts and figures regarding PCOS are limited and unclear.

● Predicting PCOS in women by using various important affecting factors and spreading awareness about the condition so that more and more people can be made self aware and can be saved from misdiagnosis.
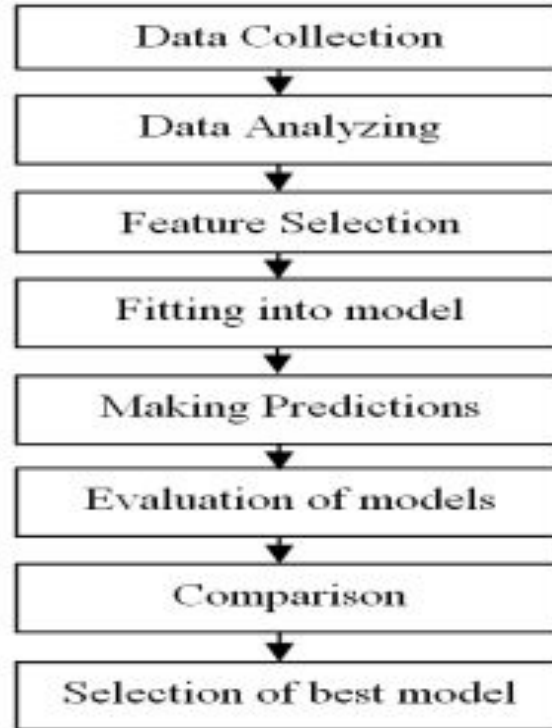
Solution description

Machine learning models are built in order to determine the presence of PCOS. Since the dataset does classify if the condition is present or not, supervised machine learning algorithms are used called: K-Nearest Neighbor (K-NN) and Logistic Regression. The former is a distance based technique and the latter is probability based, this is why these two techniques, that are poles apart, are used and their accuracy is compared

Implementation of Solution

In order to build an appropriate machine learning model, the data belonging a dataset needs to go through a sequence of steps. It is necessary, as to become a filtered and noise-less input to the algorithm. The results of the algorithms are compared and evaluated, leading to the motive, that is, detecting PCOS using machine learning model with the highest accuracy.

Workflow

| Data Collection |
| Data Analyzing |
| Feature Selection |
| Fitting into model |
| Making Predictions |
| Evaluation of models |
| Comparison |
| Selection of best model |

For PCOS prediction: we will use ML to predict if a women is healthy or not

For this we will use Azure AI Fundamentals for ML and Azure ML studio along with Azure storage vault and auto ML features of Azure..

Dataset will be from Kaggle.

For spreading Awareness: we are planning to create a bot which will answers all related questions. Which will be available in various regional languages.

For this we will use Language Understanding service from Azure NLP workload.

Knowledge base will be created using various faq documents from certified websites and services.