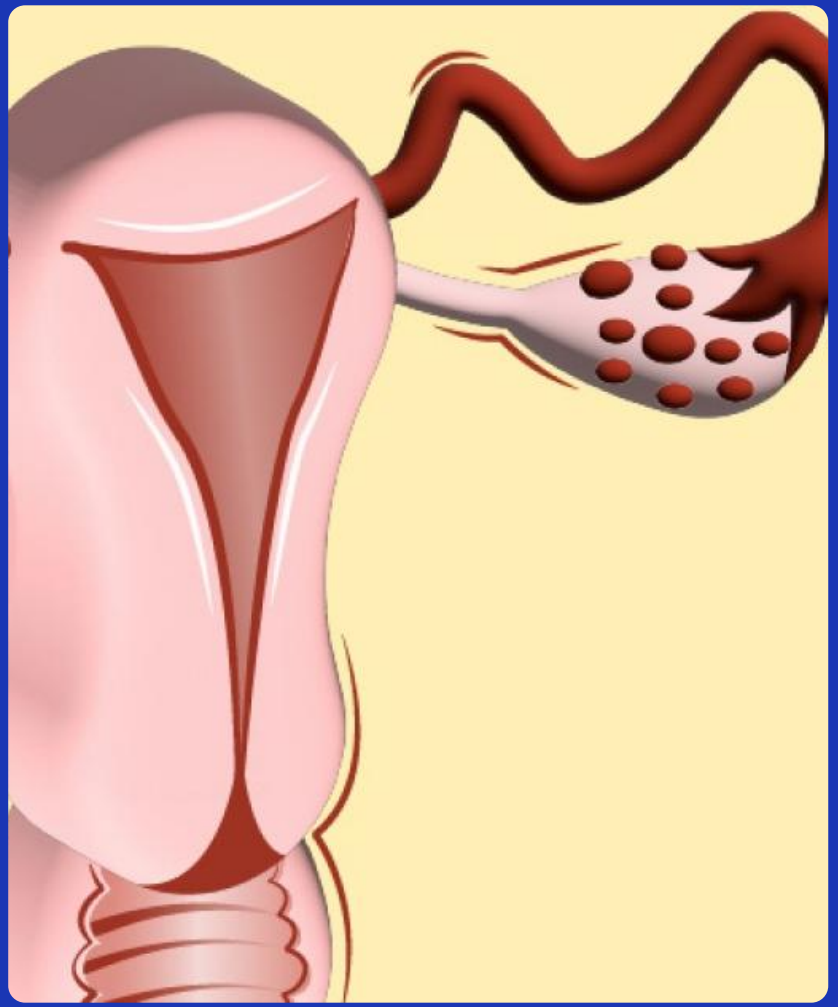
Experiment Projects

Comparative Analysis of Machine Learning Models for PCOS Diagnosis

Ken Nitipat Wuttisasiwat
Eve Saowaluk Jirapornsirikul



Approach

Data Collection

Kaggle / CSV

Fetch the PCOS dataset from Kaggle to be csv importable format

Training

sklearn

Train the model with models
- Logistic Regression
- Random Forest
- Decision Tree

Evaluation

cross_val_score

Use 5-fold validation and measure the scoring of accuracy and precision (Confusion Matrix)

Dataset

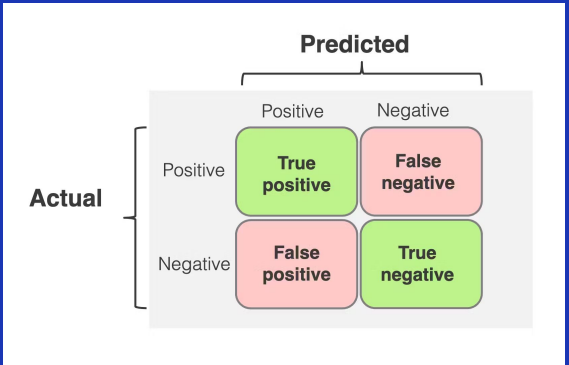
- Age (Integer)
- BMI (Decimal)
- Menstrual Irregularity (Boolean)
- Testosterone Level (Decimal)
- Antral Folliecie Count (Integer)
- PCOS Diagnosis (Boolean)

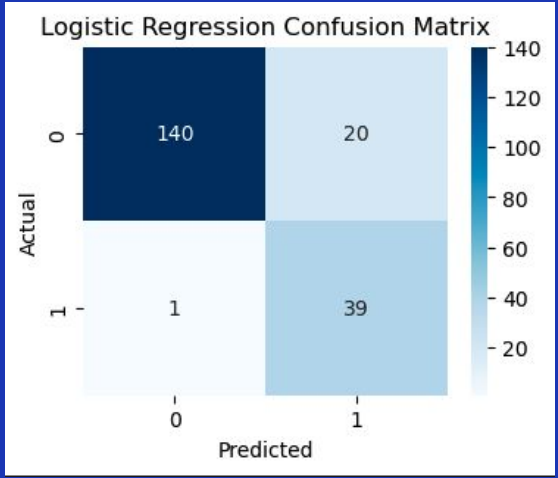
https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-pcos

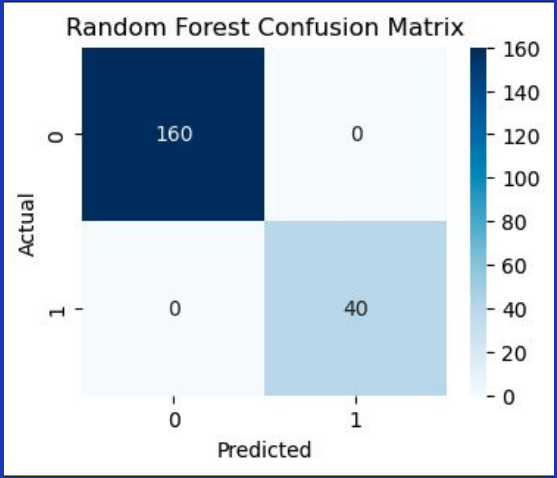
Age	BMI	Menstrual_Irregularity	Testosterone_Level(ng/dL)	Antral_Follicle_Count	PCOS_Diagnosis
24	34.7	1	25.2	20	0
37	26.4	0	57.1	25	0
32	23.6	0	92.7	28	0
28	28.8	0	63.1	26	0
25	22.1	1	59.8	8	0
38	19.3	0	28.4	6	0
24	20.2	1	72.5	29	0
43	20.2	1	85.8	17	0
36	20.6	0	50.4	5	0
40	20.4	0	82.0	21	0
28	28.9	0	97.2	17	0
28	21.1	1	36.3	8	0
41	23.9	0	61.9	11	0
38	33.2	1	43.0	21	1
21	26.1	1	83.4	13	1
25	29.3	0	66.2	8	0
41	20.9	0	70.8	23	0
20	21.3	0	83.8	5	0
39	18.7	0	51.7	14	0
38	20.9	0	93.2	8	0
19	22.7	0	62.6	13	0
41	21.0	1	32.6	17	0
29	19.5	1	75.7	18	0
23	20.1	1	83.5	14	0
19	25.8	0	45.3	10	0
45	21.5	1	88.6	20	0
38	24.2	1	92.5	10	0
18	26.6	1	42.2	21	1

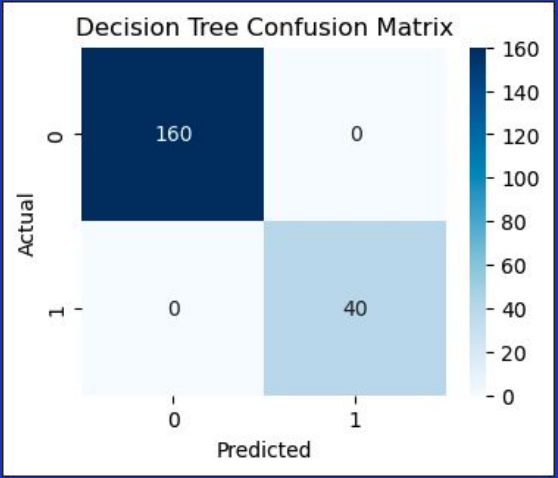
Evaluation

Confusion Matrix









Result

Random Forest and Decision Tree perform very well on the dataset

Meet the Team

Ken Nitipat Wuttisasiwat

- Explored Dataset
- Models implementation
- Presentation
- Project Report

Eve Saowaluk Jirapornsirikul

- Visualization
- Evaluation
- Presentation
- Project Report