Title: **Dementia Prediction Using ML and AutoML**

Team Members: Vishwashree V Karhadkar and Gagandeep Singh

Dataset Description: Patient health metrics dataset fetched from kaggle for dementia prediction, having a total sample count of 1000 columns covering diverse features including lifestyle, health indicators, and genetic factors.

Type of Classification: Binary classification if the person has dementia or not if yes =1 if not =0

Main Predictive Task: Standard task is to classify and predict based on dataset whether person has dementia or not.

Questions and Approach

Hypotheses/Questions

Can ML models accurately predict dementia from patient data?

Does gender-based dataset division (Male/Female) improve prediction accuracy?

What impact does removing certain columns have on model performance?

2 | Approach: Run for Phase 3

1. Run full Dataset
2. Male Dataset
3. Female Dataset

3 | Approach: Run for Phase 4

With all Features available:
1. Run full Dataset
2. Male Dataset
3. Female Dataset

Removed Prescription, Dosage in mg, Cognitive_Test_Scores
1. Full Dataset
2. Male Dataset
3. Female Dataset

Questions and Approach

Top Models:

Accuracy with Top Models

Full Dataset: sgd with 75.6%

Male Dataset: lgbm with 76.2%

Female Dataset: lgbm with 75.8%

2 | Accuracy Insights:

With all Features:

Full Dataset: 100%

Male Dataset: 100%

Female Dataset: 100%

With removed Columns:

Full Dataset: 75.6%

Male Dataset: 76.2%

Female Dataset: 75.8%

3 | Feature Impact:

Full Dataset:
['Depression_Status_Yes', 'Gender_Male', 'BodyTemperature', 'Weight', 'Age', 'HeartRate', 'BloodOxygenLevel']

Male Dataset:
['Depression_Status_Yes', 'Education_LEvel_Primary School', 'Weight', 'BodyTempreture,' 'Age', 'HeartRate','BloodOxygenLevel']

Female Dataset:
['Depression_Status_Yes', 'Diabetic_1', 'BodyTempreture Age']

DF Analyze provided insights with different model accuracy based on feature being present and not presented across dataset variations, helping identify optimized model approaches for dementia prediction.