PROJECT: MULTIPLE DISEASE PREDICTION

APPROACH:

- Cleaned data by looking for null, missing values and unknown values like Nan
- Identified the presence of any ordinal and categorical variables
- Performed ordinal encoding at required places to aid the model for better data handling
- More missing values were found in kidney disease data set and are filled appropriately using KNN imputation
- Performed Exploratory Data Analysis (EDA) to identify the correlations and key factors influencing the target
- Selected key features related to the target variable for model building based on the domain knowledge
- Checked for data balance and under sampling technique was executed to improve the accuracy of the ML model
- Chose appropriate ML algorithm for each disease using lazy predict
- Built KNN Classifier Model to predict Parkinson's disease and trained using training data set
- Built Decision Tree Classifier Model to predict Chronic Kidney disease and trained using training data set
- Built Random Forest Classifier Model to predict Liver disease and trained using training data set
- Evaluated the models using test data set and the results are shared below
- Saved the model using pickle
- Loaded models in app.py file and designed Streamlit app that enables healthcare providers to enter the values (from test results) of the features and predict the likelihood of the disease using probability
- Deployment steps has been shared below
- Inference and Suggestions for health care providers (for each disease) has been shared in the Streamlit app

Model Evaluation:

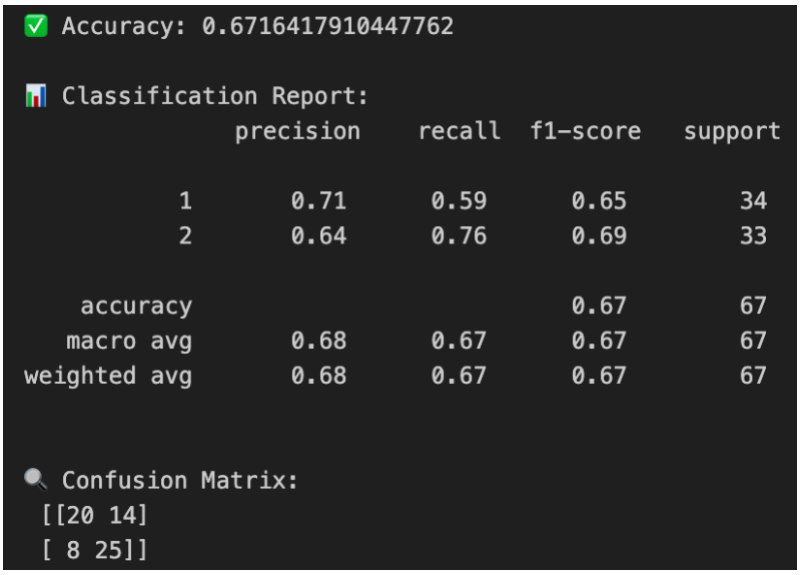
1. Parkinson's Disease – KNN Classifier:

```
Class distribution after undersampling:
status
0    48
1    48
Name: count, dtype: int64
Accuracy: 0.85

Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.90      0.86        10
           1       0.89      0.80      0.84        10

    accuracy                           0.85        20
   macro avg       0.85      0.85      0.85        20
weighted avg       0.85      0.85      0.85        20


Confusion Matrix:
 [[9 1]
 [2 8]]
```

2. Chronic Kidney Disease – Decision Tree Classifier:

```
Class distribution after undersampling:
classification
0.0    150
1.0    150
Name: count, dtype: int64
✅ Accuracy: 0.9833333333333333

📊 Classification Report:
              precision    recall  f1-score   support

         0.0       1.00      0.97      0.98        30
         1.0       0.97      1.00      0.98        30

    accuracy                           0.98        60
   macro avg       0.98      0.98      0.98        60
weighted avg       0.98      0.98      0.98        60


🔎 Confusion Matrix:
 [[29  1]
 [ 0 30]]
```

3. Liver Disease – Random Forest Classifier:



Streamlit App Deployment Instructions:

- Open a new notebook in VS code or in Google colab
- Upload the given pickle file of each model in the same folder where the notebook is saved
- Run the cells to Install Streamlit and app.py in the notebook
- Use or click the link given in the terminal; Streamlit dashboard appears in the web browser
- Select the disease(model) of your choice; Respective features appear with values of healthy individual
- Enter the values of the individual from the test results in the respective features and click Predict to see the outcome