**WEEK 1: Disease Prediction Using Patient Data**

**Objective:**

Learn basic ML workflows to predict diseases like diabetes or heart disease.

1. **Dataset:**

* Used Diabetes and Heart Disease datasets in CSV format.
* Source: UCI Machine Learning Repository.

1. **Preprocessing:**

* Missing values filled with mean for numeric columns.
* Numeric columns normalized between 0 and 1 using MinMaxScaler.
* Categorical features encoded using LabelEncoder (Gender, Yes/No features).

1. **Exploratory Data Analysis (EDA):**

* Dataset shape and column information viewed.
* Numeric statistics summarized using describe().
* Correlation heatmap plotted to understand feature relationships.

1. **Model Training:**

* Logistic Regression and Random Forest models trained on datasets.
* Train/Test split: 80% training, 20% testing.

1. **Evaluation:**

* Accuracy used as primary metric.
* Random Forest achieved higher accuracy than Logistic Regression.
* Best model selected automatically based on performance.

1. **Outcome:**

* System accepts patient input (numeric and categorical features).
* Generates disease prediction using the best-performing model.
* Diagnosis presented in user-friendly format.

**Result Example:**

* Logistic Regression Accuracy**: 0.933**
* Random Forest Accuracy**: 0.990**

Selected Model: Random Forest - Prediction for patient input: Positive / Negative (with explanation).

**Conclusion:** This workflow demonstrates a simple ML pipeline for disease prediction. - Preprocessing, model training, and evaluation steps are covered. - User can predict disease risk interactively.