**PROJECT: Multiple Disease Prediction**

**Description:** This application predicts the likelihood of multiple chronic diseases—Liver Disease, Kidney Disease, and Parkinson’s Disease—using patient medical data such as biochemical test results, clinical indicators, and physiological measurements. By identifying early risk patterns and disease-specific markers, the system supports early diagnosis, timely medical intervention, and improved healthcare decision-making.

**Technologies Used:** Python, Streamlit, ML, EDA

**Project Details:**

**Data Preparation & Pre-processing**

* Collected and analyzed **separate datasets** for:
  + Liver Disease
  + Kidney Disease
  + Parkinson’s Disease
* Cleaned datasets by:
  + Handling missing and inconsistent values
  + Removing irrelevant identifiers (e.g., patient ID, name)
* Applied:
  + **Label Encoding** for categorical medical features
  + **Numeric type conversion** for clinical attributes
  + **Log Transformation** to reduce skewness and manage outliers
* Addressed class imbalance using:
  + **SMOTE / SMOTETomek**
* Ensured consistent feature alignment between training and prediction pipelines.

**EDA**

* Performed comprehensive EDA using **Pandas and Seaborn** to:
* Understand feature distributions and correlations
* Detect outliers and skewed medical indicators
* Identify disease-specific risk factors
* Analysed relationships between:
* Lab test values and disease outcomes
* Clinical indicators and diagnosis results

**Visualization**

* Built an interactive Streamlit dashboard for HR stakeholders.
* Visualized:
  + Disease distribution across patients
  + Boxplots highlighting abnormal clinical ranges
  + Correlation heatmaps of medical features
  + Comparative analysis of healthy vs diseased cases

**Machine Learning Models**

* Trained and evaluated multiple models for **Attrition Prediction and Performance Prediction**:
  + Logistic Regression
  + Random Forest
  + Decision Tree
  + K-Nearest Neighbors (KNN)
  + Support Vector Machine (SVM)
  + AdaBoost
  + Gradient Boosting
  + XGBoost

**Model Evaluation Metrics:**

* Accuracy
* Precision
* Recall
* F1-Score
* AUC-ROC

**Model Selection & Deployment**

* Selected the best-performing model based on evaluation metrics.
* Saved the trained model using Pickle.
* Integrated the model into Streamlit for:
  + Liver Disease Prediction
  + Kidney Disease Prediction
  + Parkinson’s Disease Prediction

**Python Code:**







**Pickle files:**



**Screenshots**













