Logistic Regression

**1. Data Exploration:**

a. Load the dataset and perform exploratory data analysis (EDA).

b. Examine the features, their types, and summary statistics.

c. Create visualizations such as histograms, box plots, or pair plots to visualize the distributions and relationships between features.

Analyze any patterns or correlations observed in the data.

**2. Data Preprocessing:**

a. Handle missing values (e.g., imputation).

b. Encode categorical variables.

**3. Model Building:**

a. Build a logistic regression model using appropriate libraries (e.g., scikit-learn).

b. Train the model using the training data.

**4. Model Evaluation:**

a. Evaluate the performance of the model on the testing data using accuracy, precision, recall, F1-score, and ROC-AUC score.

Visualize the ROC curve.

**5. Interpretation:**

a. Interpret the coefficients of the logistic regression model.

b. Discuss the significance of features in predicting the target variable (survival probability in this case).

**6. Deployment with Streamlit:**

In this task, you will deploy your logistic regression model using Streamlit. The deployment can be done locally or online via Streamlit Share. Your task includes creating a Streamlit app in Python that involves loading your trained model and setting up user inputs for predictions.

(optional)For online deployment, use Streamlit Community Cloud, which supports deployment from GitHub repositories.

Detailed deployment instructions are available in the Streamlit Documentation.

<https://docs.streamlit.io/streamlit-community-cloud/deploy-your-app>

**Interview Questions:**

1. What is the difference between precision and recall?

2. What is cross-validation, and why is it important in binary classification?