Classification Task

Dataset Link: Kaggle Dataset

Task Overview:

- Data Cleaning:
  - Perform Exploratory Data Analysis (EDA) to understand the dataset.
  - Identify and handle missing values, outliers, or any other data quality issues.
  - Choose appropriate techniques to handle categorical variables if needed.
- Classification Models:
  - Implement the following classification models using the cleaned dataset:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree
  - Random Forest
- Hyperparameter Tuning:
  - Apply hyperparameter tuning for each of the classification models to optimize their performance.
  - Techniques like Grid Search or Random Search can be used.
  - Document the hyperparameters tuned and the rationale behind the choices.
- Model Comparison:
  - Train each model on the dataset and evaluate their performance using classification metrics (e.g., accuracy, precision, recall, F1-score).
  - Provide a detailed comparison across models, discussing their strengths, weaknesses, and any interesting observations.
  - Create visualizations (e.g., confusion matrices, ROC curves) to enhance the comparison.