Capstone Project: Machine Learning for Classification and Regression

Objective:
The objective of this capstone project is to develop a comprehensive machine learning solution for classification and regression tasks. The project will involve various stages including data collection, preprocessing, model training, evaluation, hyperparameter tuning, and deployment of the best-performing model as a web application.

Dataset:
- The dataset for this project should contain substantial noise and encompasses at least 15 features.
- Data collection will involve fetching the dataset and uploading it to a SQL Server database for storage and management.

Project Workflow:
1. Data Collection and Preprocessing:
   - Fetch the dataset.
   - Upload the data to a SQL Server database.
   - Preprocess the data to handle missing values, encode categorical variables, and scale numerical features.

2. Exploratory Data Analysis (EDA):
   - Perform EDA to gain insights into the data distribution, relationships between variables, and potential patterns.
   - Utilize summary statistics, histograms, scatter plots, and correlation matrices for exploration.

3. Modeling:
   - Apply various machine learning algorithms suitable for classification and regression tasks.
   - Evaluate the performance of each model using appropriate metrics.
   - Perform hyperparameter tuning to optimize model performance.

4. Model Evaluation:
   - Select the best-performing model based on evaluation metrics.
   - Justify the selection of the best model.

5. Web Application Development:
   - Develop a web application to deploy the best-performing model.
   - Provide user-friendly interfaces for input and display of results.

6. Reporting:
   - Prepare a detailed data analysis and modeling report.
   - Create a PowerPoint presentation summarizing the project workflow, findings, and conclusions.

Deliverables:
- Completed Jupyter Notebook or Python script containing all code and analysis.
- SQL script for data uploading and manipulation.
- Web application URL and files for model deployment.
- Detailed data analysis and modeling report.
- PowerPoint presentation slides.

Evaluation Criteria:
- Completion of each project stage.
- Quality of data preprocessing and exploratory data analysis.
- Effectiveness of machine learning models and model evaluation.
- Successful deployment of the web application.
- Clarity and coherence of the data analysis and modeling report.

Resources:
- Sample Dataset Source: [https://www.kaggle.com/]