**Iris Classification Project Documentation**

**1. Project Overview**

* **Title:** Iris Classification
* **Objective:** To classify iris flower species based on their features using machine learning algorithms.
* **Dataset:** Iris dataset, available from the UCI Machine Learning Repository.
* **Features:** Sepal length, sepal width, petal length, petal width.
* **Classes:** Iris-setosa, Iris-versicolor, Iris-virginica.

**2. Data Description**

* **Source:** UCI Machine Learning Repository
* **Attributes:**
  + Sepal Length (cm)
  + Sepal Width (cm)
  + Petal Length (cm)
  + Petal Width (cm)
* **Target Variable:** Species (Iris-setosa, Iris-versicolor, Iris-virginica)
* **Number of Instances:** 150
* **Number of Features:** 4

**3. Data Preprocessing**

* **Handling Missing Values:** None (Iris dataset is complete)
* **Feature Scaling:** Standardization (if applicable)
* **Splitting the Dataset:** Training set and test set (e.g., 80% training, 20% testing)

**4. Exploratory Data Analysis (EDA)**

* **Visualization:** Histograms, pair plots, scatter plots
* **Statistical Summary:** Mean, median, standard deviation of features
* **Correlation Analysis:** Heatmap or correlation matrix

**5. Model Selection**

* **Algorithms Used:** (e.g., Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Decision Trees)
* **Evaluation Metrics:** Accuracy, Precision, Recall, F1 Score

**6. Model Training**

* **Training Process:** Description of the training process for each model
* **Hyperparameter Tuning:** Grid search, cross-validation (if applicable)

**7. Results**

* **Model Performance:** Evaluation metrics for each model
* **Confusion Matrix:** For each model
* **Comparison:** Summary of which model performed best

**8. Conclusion**

* **Summary of Findings:** Insights gained from the model performance
* **Future Work:** Potential improvements or additional experiments

**9. References**

* **Dataset:** UCI Machine Learning Repository
* **Libraries Used:** scikit-learn, pandas, matplotlib, seaborn

**Wine Quality Classification Project Documentation**

**1. Project Overview**

* **Title:** Wine Quality Classification
* **Objective:** To predict the quality of wine based on its physicochemical properties using machine learning algorithms.
* **Dataset:** Wine Quality dataset, available from the UCI Machine Learning Repository.
* **Features:** Fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol.
* **Target Variable:** Quality (integer scale from 0 to 10)

**2. Data Description**

* **Source:** UCI Machine Learning Repository
* **Attributes:**
  + Fixed Acidity
  + Volatile Acidity
  + Citric Acid
  + Residual Sugar
  + Chlorides
  + Free Sulfur Dioxide
  + Total Sulfur Dioxide
  + Density
  + pH
  + Sulphates
  + Alcohol
* **Target Variable:** Quality (integer values)
* **Number of Instances:** 4898 (for red wine); 1599 (for white wine)
* **Number of Features:** 11

**3. Data Preprocessing**

* **Handling Missing Values:** Imputation or removal
* **Feature Scaling:** Normalization or Standardization
* **Splitting the Dataset:** Training set and test set (e.g., 70% training, 30% testing)

**4. Exploratory Data Analysis (EDA)**

* **Visualization:** Histograms, box plots, pair plots, correlation heatmap
* **Statistical Summary:** Mean, median, standard deviation of features
* **Correlation Analysis:** Insights into which features are most correlated with quality

**5. Model Selection**

* **Algorithms Used:** (e.g., Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine)
* **Evaluation Metrics:** Accuracy, Precision, Recall, F1 Score, Confusion Matrix

**6. Model Training**

* **Training Process:** Description of the training process for each model
* **Hyperparameter Tuning:** Grid search, cross-validation (if applicable)

**7. Results**

* **Model Performance:** Evaluation metrics for each model
* **Confusion Matrix:** For each model
* **Comparison:** Summary of which model performed best

**8. Conclusion**

* **Summary of Findings:** Insights gained from the model performance
* **Future Work:** Potential improvements or additional experiments

**9. References**

* **Dataset:** UCI Machine Learning Repository
* **Libraries Used:** scikit-learn, pandas, matplotlib, seaborn