1. Theory and Definitions:

Iris Dataset:
The Iris flower dataset is one of the most famous datasets in the field of machine learning and statistics. It consists of 150 instances of Iris flowers, divided into three species: setosa, versicolor, and virginica. Each instance has 4 features:

- **Sepal Length: The length of the sepal of the flower (numeric).**

- **Sepal Width: The width of the sepal (numeric).**

- **Petal Length: The length of the petal (numeric).**

- **Petal Width: The width of the petal (numeric).**

- **Species: The species of the Iris flower (nominal: setosa, versicolor, virginica).**

2. Algorithm (Steps):

1. **Load Dataset: Import the Iris dataset using seaborn and load it into a pandas DataFrame.**

2. **Data Inspection: View the dataset's first few rows and check the data types of each feature using dataset.head() and dataset.info().**

3. **Histograms:**

   - **Create a set of histograms to visualize the distribution of each numeric feature (sepal length, sepal width, petal length, petal width).**

   - **This helps understand how the data is spread and whether there are any skewed distributions.**

4. **Boxplots:**

   - **Create boxplots for each feature (sepal length, sepal width, petal length, petal width) grouped by species. This helps in visualizing the median, quartiles, and any outliers for each feature.**

5. **Outlier Detection:**

    ○ **Outliers are detected using the boxplot, as any data point outside the interquartile range (IQR) is considered an outlier.**

6. **Visualization: Use `matplotlib` and `seaborn` to visualize histograms and boxplots in subplots.**

3. Observations and Inferences:

- **Features:**

    ○ **The features in the Iris dataset include numeric values such as `sepal_length`, `sepal_width`, `petal_length`, and `petal_width`, while the target variable `species` is nominal.**

- **Histogram Analysis:**

    ○ **`Sepal length` is relatively symmetrically distributed, with a peak around 5.0 to 6.0.**

    ○ **`Petal length` and `width` show more spread for the "virginica" species, while the "setosa" species has smaller values for both features.**

    ○ **`Sepal width` shows a slight skew toward the right.**

- **Boxplot Analysis:**

    ○ **The boxplots confirm the differences in the central tendency and spread between the species.**

    ○ **"Virginica" flowers have the largest petal dimensions, while "setosa" has the smallest.**

- **Outliers:**

    ○ **The boxplots show a few outliers in "virginica," particularly for `petal_length` and `petal_width`.**

4. Conclusion:

The Iris dataset provides valuable insight into the distribution and characteristics of Iris flowers' features. Histograms and boxplots offer a clear visualization of the distribution and differences across the species. "Setosa" flowers have distinctively smaller sepal and petal sizes compared to the other species, while "Virginica" shows larger petal sizes with a few outliers. The dataset is well-suited for classification tasks in machine learning, and the visualizations help in understanding the spread and variability of each feature.