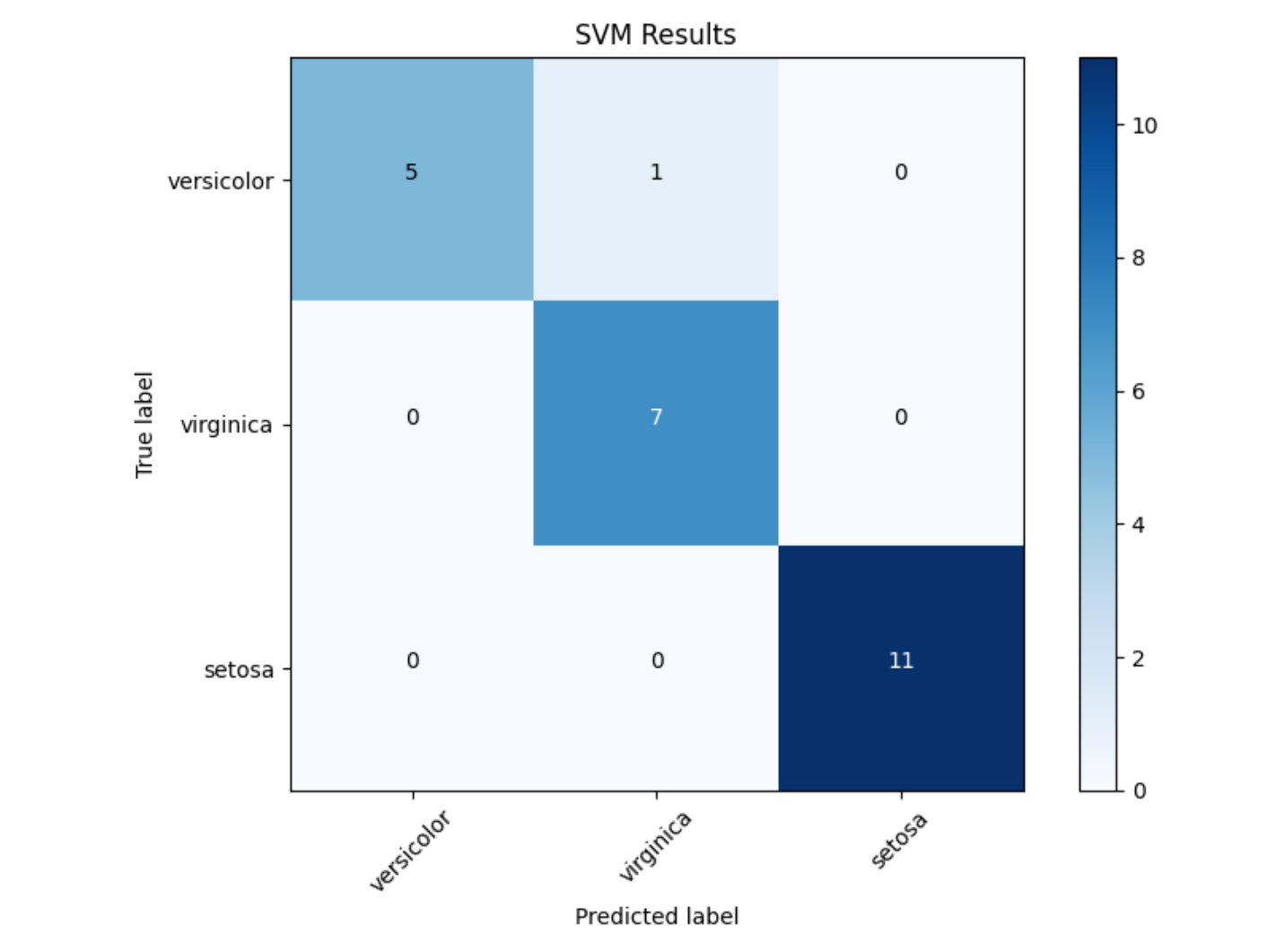
1)

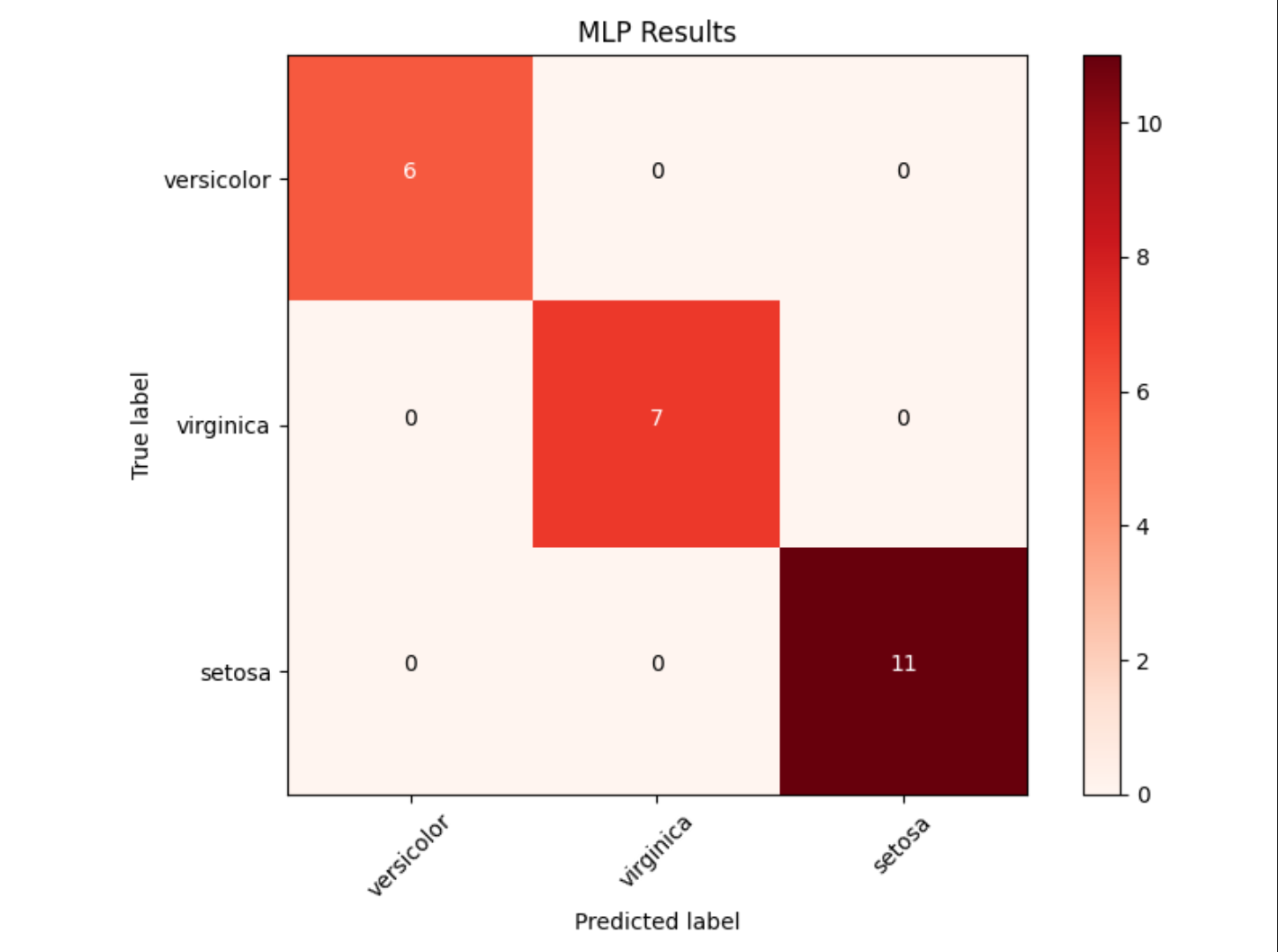
1. **Load the Iris Dataset**:
   * Download the dataset from the provided GitHub link.
   * Load the data into a DataFrame.
2. **Preprocess the Data**:
   * Combines 4 flower measurements into a feature vector
   * Normalizes features for better model convergence
   * Converts species labels to numerical indices
   * Split the data into 80% for training and 20% for testing
3. **Train and Evaluate Models**:
   * **Linear SVM**: Use LinearSVC from PySpark MLlib.
   * **Multilayer Perceptron Classifier**: Use MultilayerPerceptronClassifier and define the network layers.
   * Both models use the same preprocessing pipeline
   * Accuracy calculated using MulticlassClassificationEvaluator
4. **Visualization**:
   * Use Python's Matplotlib to visualize classification results











2)

1. **Load the Dataset**:
   * Load the dataset in LIBSVM format.
2. **Train K-Means Model**:
   * Import KMeans from pyspark.ml.clustering
   * Configures K-Means with k=2
   * Trains the model on the dataset.
3. **Output Results**:
   * Extracts cluster centers as NumPy arrays.
   * Prints centers to terminal
4. **Visualization**:
   * **Cluster 1**: Red x marks
   * **Cluster 2**: Blue o circles
   * **Centers**: Triangles (red) and squares (blue)





