Pseudocodice: Decision Tree

```
CLASS DecisionNode:
    feature     // indice feature di split
    threshold   // valore della feature su cui effettuare split
    left        // figlio sinistro
    right       // figlio destro
    value       // etichetta di classe per i nodi foglia, altrimenti null
```

```
FUNCTION BuildTree(dataset, labels)
    IF all labels are equal
        RETURN leaf with that label

    FOR each feature
        FOR each possible split value
            compute Gini impurity

    SELECT split with minimum impurity

    CREATE left and right subsets

    left_subtree  ← BuildTree(left subset)
    right_subtree ← BuildTree(right subset)

    RETURN decision node
```

```
FUNCTION gini_impurity(y)
    FOR each class
        calculates impurity

    RETURN impurity // grado di disordine dei dati
```

```
FUNCTION decision_tree_predict(tree, sample)
    IF node is a leaf node
        return node value

    IF sample's feature value is <= than the node's threashold
        return decision_tree_predict(left tree, sample)
    ELSE
        return decision_tree_predict(right tree, sample)
```

```
FUNCTION cross_validation_score(X, y, fold number)
    FOR each fold
        train the decison tree, make predictions and save the accuracy

    RETURN accuracies mean
```

```
FUNCTION load_and_preprocess_data_mushrooms()
    transform the class column with LabelEncoder // valori 0 e 1 per
commestibile e velenoso
    remove veil-type and stalk-root columns due to poor impact
    convert the attributes with one-hot-encoding and scale the information

    RETURN attributes and classes
```

```
FUNCTION load_and_preprocess_data_rice()
    convert the input data into utf-8 strings
    transform the class colum with LabelEncoder
    scale the features

    RETURN attributes and classes
```

```
FUNCTION calculate_metrics(predictions, y_true)
    RETURN accuracy, precision and recall

FUNCTION evaluate_model(tree, X_test, y_test)
    calls the functions to predict data and evaluate the model_selection
    and then prints the results

FUNCTION main()
    The main enables the user to choose one of the datasets to train and test
the decision tree,
    then calls all the functions and prints all the results and statistics
```