



Mushroom Classification: Can I Eat That??

Classifying mushrooms based on their physical characteristics is possible, but training is needed.

By: Sharon Colson, Tania Perdomo Flores, Edward Gannod

Dataset Dimension and Visualization

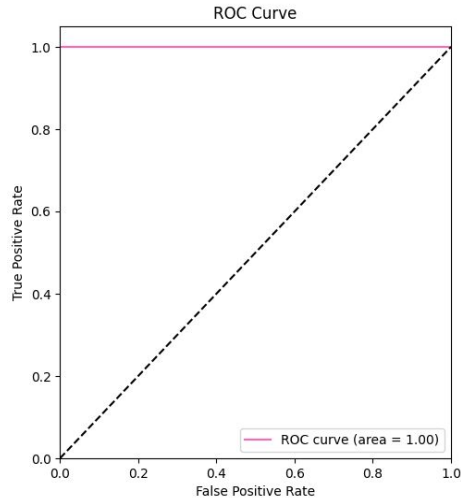
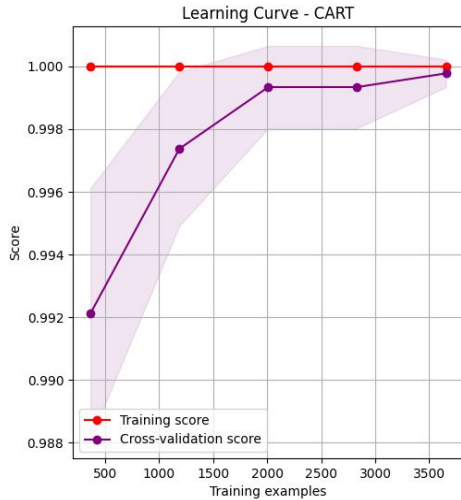
- Shape: (8124,23)
- Labels: {0:Poisonous, 1:Edible}
- Encoding: Label Encoding, One-Hot Encoding
- Scaling: Not needed.
- Feature Example: Cap Description, Bruises, Odor, Gill and Stock, Habitat.

Create an Array of Algorithms

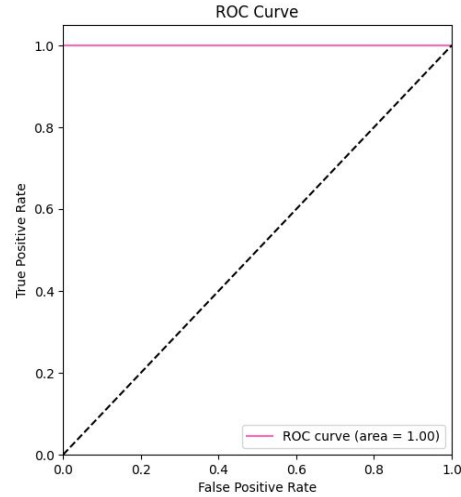
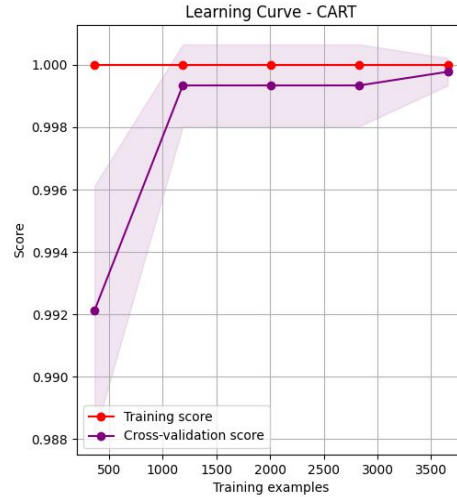
```
] # Initializes an array called models.  
models = []  
models.append(('LR', LogisticRegression()))  
models.append(('LDA', LinearDiscriminantAnalysis()))  
models.append(('KNN', KNeighborsClassifier()))  
models.append(('CART', DecisionTreeClassifier()))  
models.append(('FOREST', RandomForestClassifier()))  
models.append(('NB', GaussianNB()))  
models.append(('SVM', SVC()))  
  
] for name, model in models:  
    evaluate_model(model, X_train, y_train, X_test, y_test, name)
```

Decision Tree Classifier

CART - Accuracy: 1.0, F1 Score: 1.0



CART - Accuracy: 1.0, F1 Score: 1.0

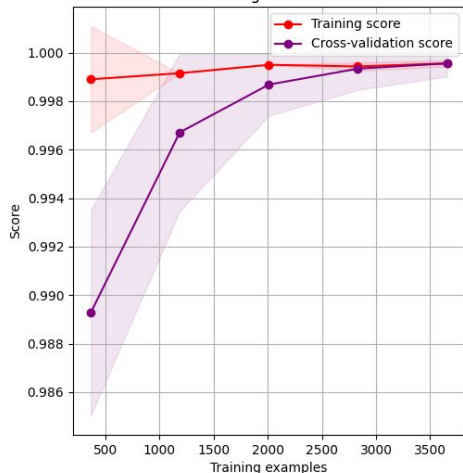


Best parameters for Decision Tree Classifier (Randomized)
'min_samples_split': 5, 'min_samples_leaf': 1, 'max_depth': 20

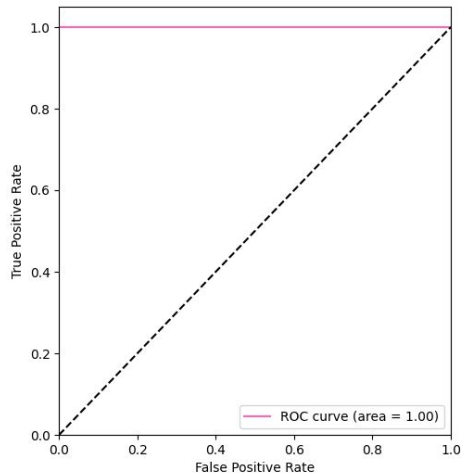
Logistic Regression

LR - Accuracy: 0.9975381585425899, F1 Score: 0.9975550122249389

Learning Curve - LR

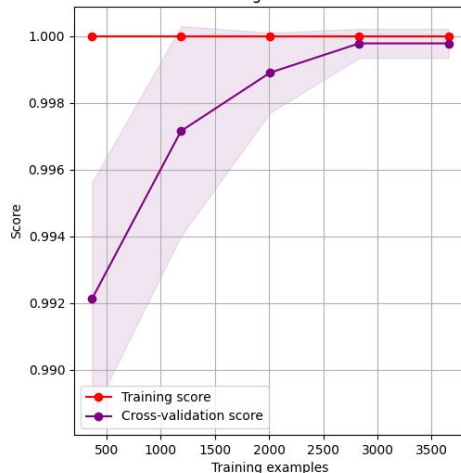


ROC Curve

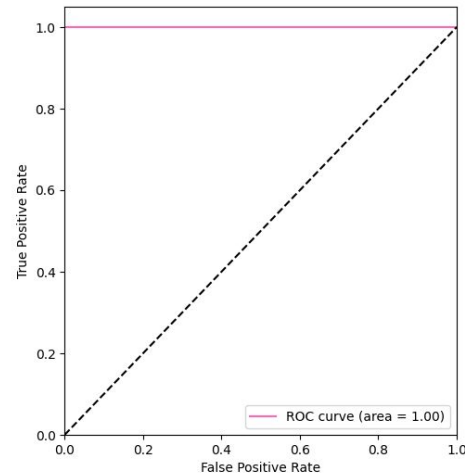


LR - Accuracy: 0.999507631708518, F1 Score: 0.9995100440960314

Learning Curve - LR



ROC Curve



Best parameters for Logistic Regression (Randomized)
'solver': 'lbfgs', 'penalty': 'l2', 'C': 10.0

Additional Information

- Hyper-Tuning: Randomized Search CV, Iterations
- Poor performing model: Naive Bayes
- Other highly performing models:
 - Linear Discriminant Analysis
 - K Neighbors Classifier (KNN)
 - Support Vector Classifier (SVC)
 - Random Forest Classifier
- Evaluation Metrics: Accuracy, F1 Score, Learning Curve ROC, Confusion Matrix, Classification Report.

▼ Parameter Grids

```
[112] # Logistic Regression
      param_lr = {
          'C': [0.1, 1.0, 10.0],
          #'penalty': ['l1', 'l2']
          'penalty': ['l2'],
          'solver': ['lbfgs', 'liblinear', 'saga']
      }

      # Decision Tree Classifier
      param_dt = {
          'max_depth': [None, 10, 20, 30],
          'min_samples_split': [2, 5, 10],
          'min_samples_leaf': [1, 2, 4]
      }

      # Random Forest Classifier
      param_rf = {
          'n_estimators': [100, 200, 300],
          'max_depth': [None, 10, 20, 30],
          'min_samples_split': [2, 5, 10],
          'min_samples_leaf': [1, 2, 4]
      }
```




Questions?

The background is a solid light pink color. It is decorated with several stylized, colorful elements: a large red flower with a yellow center at the top center; a white flower with a dark red center at the top center; a large yellow flower with a pink center at the bottom center; a white flower with a dark red center at the bottom center; and several purple umbrellas with yellow canopies and pink handles, some of which are open and some are closed. The text "Thank you!" is written in a large, white, rounded font with a pink outline, centered on the page.

Thank you!