# Data Science: A programming approach

# Group 8 - Orange

Andrew Kaiser (1000649138), Sahej Vijan (1001678867), Reddy Sushmitha Yethirajula (1001716026), Abhishek Bussa (1001624753)

## Mini Project 2. Image Analysis

Dataset:

We used images of fruits (Banana, Pineapple, Kiwi, Apple). Each fruit has 20 images each stored in ‘images’ folder.

### **Neural Networks**:

Accuracy, Precision, Recall, F-1 scores for Simple train test method

```
In-sample Accuracy:  91.07142857142857
Out-of-sample Accuracy:  58.333333333333336
             precision    recall  f1-score   support

     Banana       0.43      0.50      0.46         6
  Pineapple       0.62      0.83      0.71         6
       Kiwi       0.67      0.33      0.44         6
      Apple       0.67      0.67      0.67         6

avg / total       0.60      0.58      0.57        24
```

Accuracy, Precision, Recall, F-1 scores using Stratified Cross Validation

```
Cross Validation Scores for Neural Networks: [0.4375 0.25   0.6875 0.25
0.5625] with highest being 0.6875
Mean Accuracy Score for Neural Networks: 0.4375
As we can see, the accuracy changed from 58.333333333333336 without Strati
fied CV to a mean accuracy of 0.4375 when using Stratified CV with a max a
ccuracy of 0.6875
```

Best Parameters using Grid Search: `{'hidden_layer_sizes': 7}`
Accuracy, Precision, Recall, F-1 scores using Grid Search with Stratified Cross Validation

```
In-sample Accuracy:  91.07142857142857
Out-of-sample Accuracy:  58.333333333333336
             precision    recall  f1-score   support

     Banana       0.43      0.50      0.46         6
  Pineapple       0.62      0.83      0.71         6
       Kiwi       0.67      0.33      0.44         6
      Apple       0.67      0.67      0.67         6

avg / total       0.60      0.58      0.57        24
```

**Random Forest:**
Accuracy, Precision, Recall, F-1 scores for Simple train test method

```
Out-of-sample Accuracy:  66.66666666666666
In-sample Accuracy:  96.42857142857143
             precision    recall  f1-score   support

     Banana       0.43      0.50      0.46         6
  Pineapple       0.40      0.33      0.36         6
       Kiwi       0.29      0.33      0.31         6
      Apple       0.60      0.50      0.55         6

avg / total       0.43      0.42      0.42        24
```

Accuracy, Precision, Recall, F-1 scores using Stratified Cross Validation

```
Cross Validation Scores for Random Forest: [0.375  0.5625 0.6875 0.3125 0.
5   ]
Mean Accuracy Score for Random Forest: 0.4875
As we can see, the accuracy changed from 41.66666666666667 without Stratif
ied CV to a mean accuracy of 0.4875 when using Stratified CV with a max ac
curacy of 0.6875
```

Best Parameters using Grid Search: `{'pca__n_components': 15, 'randomforestcla
ssifier__max_depth': 9, 'randomforestclassifier__n_estimators': 9}`

Accuracy, Precision, Recall, F-1 scores using Grid Search with Stratified Cross Validation

```
In-sample Accuracy:  98.21428571428571
Out-of-sample Accuracy:  45.83333333333333
             precision    recall  f1-score   support

     Banana       0.38      0.50      0.43         6
  Pineapple       0.60      0.50      0.55         6
       Kiwi       0.40      0.33      0.36         6
      Apple       0.50      0.50      0.50         6

avg / total       0.47      0.46      0.46        24
```

**Support Vector Machine:**

Accuracy, Precision, Recall, F-1 scores for Simple train test method

```
Out-of-sample Accuracy:  79.16666666666666
In-sample Accuracy:  89.28571428571429
             precision    recall  f1-score   support

     Banana       0.83      0.83      0.83         6
  Pineapple       1.00      1.00      1.00         6
       Kiwi       0.71      0.83      0.77         6
      Apple       0.60      0.50      0.55         6

avg / total       0.79      0.79      0.79        24
```

Accuracy, Precision, Recall, F-1 scores using Stratified Cross Validation

```
Cross Validation Scores for Svm: [0.625  0.6875 0.8125 0.4375 0.625 ]
Mean Accuracy Score for Svm: 0.6375
As we can see, the accuracy changed from 79.16666666666666 without Stratif
ied CV to a mean accuracy of 0.6375 when using Stratified CV with a max ac
curacy of 0.8125
```

Best Parameters using Grid Search: `{'svc__C': 5, 'svc__gamma': 0.005}`

Accuracy, Precision, Recall, F-1 scores using Grid Search with Stratified Cross Validation

```
Out-of-sample Accuracy:  75.0
In-sample Accuracy:  75.0
             precision    recall  f1-score   support

     Banana       0.71      0.83      0.77         6
  Pineapple       1.00      1.00      1.00         6
       Kiwi       0.75      0.50      0.60         6
      Apple       0.57      0.67      0.62         6

avg / total       0.76      0.75      0.75        24
```

As we can see, Support Vector Machines gives the highest accuracy of 79.166% when using simple test train split.