```
====================================
> > > > Import des librairies < < < <
====================================


========================================
> > > > Définition des fonctions < < < <
========================================


============================================================
> > > > TRAITEMENT DU JEU DE DONNES D'ENTRAINEMENT < < < <
============================================================


============================================================
Identification des chemins d'accès aux répertoires d'images
============================================================

Nombre d'images par catégorie (sous-répertoire):

    Catégorie  Nombre d'images
0       Corn               50
1  Raspberry               50
2     Orange               50

Nombre total d'images: 150

dataset_path = data/fruits_360_v3b/Training/

image_path =
data/fruits_360_v3b/Training/Corn,data/fruits_360_v3b/Training/Raspberry,data/fr
uits_360_v3b/Training/Orange

Nombre de catégories de fruits: 3

2 premières catégories: ['data/fruits_360_v3b/Training/Corn',
'data/fruits_360_v3b/Training/Raspberry']
2 dernières catégories: ['data/fruits_360_v3b/Training/Raspberry',
'data/fruits_360_v3b/Training/Orange']

Durée de l'opération 'Récupération des images': 0.01 s


======================
Calcul des descripteurs
======================


Chargement des images (rdd_images)
==================================

MapPartitionsRDD[4] at javaToPython at NativeMethodAccessorImpl.java:0
```

```
Nombre de partitions: 5
Dimension: 150

Catégories / Images / Descripteurs (rdd_cat_ima_desc)
=====================================================

PythonRDD[9] at RDD at PythonRDD.scala:53

Catégories / Images / Descripteurs (rdd_cat_ima_desc_f)
=======================================================

PythonRDD[10] at RDD at PythonRDD.scala:53

Catégories (rdd_cat)
====================

PythonRDD[14] at RDD at PythonRDD.scala:53

Identifiants des images (rdd_ima)
=================================

PythonRDD[15] at RDD at PythonRDD.scala:53

Descripteurs (rdd_desc)
=======================

PythonRDD[16] at RDD at PythonRDD.scala:53

Nombre de partitions: 5
Dimension: 11133

Collecte des catégories d'images (list_cat)
===========================================

3 premières occurences: ['Raspberry', 'Raspberry', 'Raspberry']

Collecte des identifiants des images (list_ima)
===============================================

3 premières occurences: ['Raspberry_19_100.jpg', 'Raspberry_19_100.jpg',
'Raspberry_19_100.jpg']
df_ima_cat: (11133, 2)
df_ima_cat (sans dup): (150, 2)

Identifiants des images et des catégories (sdf_ima_cat)
=======================================================

root
 |-- ima: string (nullable = true)
 |-- cat: string (nullable = true)

+--------------------+---------+
|                 ima|      cat|
```

```
+--------------------+---------+
|Raspberry_19_100.jpg|Raspberry|
|Raspberry_14_100.jpg|Raspberry|
|Raspberry_43_100.jpg|Raspberry|
+--------------------+---------+

Durée de l'opération 'Extraction des descripteurs des images': 18.66 s


============================================================
Classification non supervisée des descripteurs avec K-Means
============================================================


Modèle K-Means (km_model)
=========================

<pyspark.mllib.clustering.KMeansModel object at 0x7f1d88097790>

Nombre de clusters: 30

Durée de l'opération 'Clustering K-Means': 32.76 s


========================================
Prédictions des descripteurs avec K-Means
========================================


Prédictions (rdd_km_pred)
=========================

PythonRDD[226] at RDD at PythonRDD.scala:53

Nombre de partitions: 5
Dimension: 11133

Collecte des prédictions (list_km_pred)
=======================================

[3, 21, 11, 27, 7, 11, 13, 11, 20, 1]

Durée de l'opération 'Prédiction K-Means': 6.2 s


========================
Création du bag of words
========================


Encodage des identifiants d'images et concatenation avec les prédictions
(clusters K-Means)
================================================================================
===========
```

Encodage des identifiants d'images (sdf_ima_label)

+--------------------+--------+----------+
| IMA|image_id|prediction|
+--------------------+--------+----------+
|Raspberry_19_100.jpg| 110| 3|
|Raspberry_19_100.jpg| 110| 21|
|Raspberry_19_100.jpg| 110| 11|
|Raspberry_19_100.jpg| 110| 27|
|Raspberry_19_100.jpg| 110| 7|
|Raspberry_19_100.jpg| 110| 11|
|Raspberry_19_100.jpg| 110| 13|
|Raspberry_19_100.jpg| 110| 11|
|Raspberry_19_100.jpg| 110| 20|
|Raspberry_19_100.jpg| 110| 1|
+--------------------+--------+----------+

Prédictions (clusters K-Means) par image (sdf_ima_pred)
---

root
 |-- id: long (nullable = true)
 |-- prediction: long (nullable = true)

+---+----------+
| id|prediction|
+---+----------+
|110| 3|
|110| 21|
|110| 11|
|110| 27|
|110| 7|
|110| 11|
|110| 13|
|110| 11|
|110| 20|
|110| 1|
+---+----------+

Liste des clusters par image (Map + reduceByKey)
==========

Clusters par image (rdd_words)

PythonRDD[256] at RDD at PythonRDD.scala:53

Nombre de partitions: 2
Dimension: 150

Liste de 'words' par image (sdf_worcds)

```
---------------------------------------

root
 |-- image_id: long (nullable = true)
 |-- words: array (nullable = true)
 |    |-- element: string (containsNull = true)

+--------+--------------------+
|image_id|               words|
+--------+--------------------+
|     110|[3, 21, 11, 27, 7...|
|     112|[3, 3, 3, 27, 11,...|
|     108|[19, 13, 27, 3, 7...|
|     106|[3, 1, 3, 27, 27,...|
|     114|[19, 3, 7, 3, 11,...|
|     116|[21, 3, 7, 23, 21...|
|     140|[27, 3, 8, 15, 27...|
|     124|[13, 7, 0, 3, 3, ...|
|     102|[13, 3, 21, 27, 3...|
|     136|[3, 19, 27, 8, 27...|
+--------+--------------------+


Création du bag of words à partir des listes de 'words' associées aux images
(CountVectorizer)
================================================================================
==============

Bag of words (sdf_bow)
----------------------

root
 |-- image_id: long (nullable = true)
 |-- bag_of_words: vector (nullable = true)

+--------+--------------------+
|image_id|        bag_of_words|
+--------+--------------------+
|     110|(30,[0,1,2,3,4,5,...|
|     112|(30,[0,1,2,3,4,5,...|
|     108|(30,[0,1,2,3,4,5,...|
|     106|(30,[0,1,2,3,4,5,...|
|     114|(30,[0,1,2,3,4,5,...|
|     116|(30,[0,1,2,3,4,5,...|
|     140|(30,[0,1,2,3,4,5,...|
|     124|(30,[0,1,2,3,4,5,...|
|     102|(30,[0,1,2,3,4,5,...|
|     136|(30,[0,1,2,3,4,5,...|
+--------+--------------------+


==========================
Sauvegarde du bag of words
==========================
```

Bag of words (df_bow)
==========

	image_id	bag_of_words
0	110	(3.0, 9.0, 4.0, 15.0, 4.0, 9.0, 10.0, 1.0, 3.0...
1	112	(2.0, 8.0, 6.0, 13.0, 2.0, 7.0, 13.0, 0.0, 7.0...
2	108	(3.0, 12.0, 2.0, 13.0, 4.0, 8.0, 11.0, 3.0, 4....
3	106	(2.0, 10.0, 6.0, 14.0, 3.0, 8.0, 15.0, 1.0, 4....
4	114	(2.0, 8.0, 4.0, 10.0, 4.0, 7.0, 11.0, 1.0, 5.0...

Bag of words
==========

	ima	cat	0	1	2	3	4	5
0	Raspberry_19_100.jpg	Raspberry	3.0	9.0	4.0	15.0	4.0	9.0
1	Raspberry_14_100.jpg	Raspberry	2.0	8.0	4.0	12.0	3.0	10.0
2	Raspberry_43_100.jpg	Raspberry	2.0	8.0	24.0	11.0	2.0	9.0
3	Raspberry_45_100.jpg	Raspberry	2.0	7.0	18.0	9.0	2.0	9.0
4	Raspberry_20_100.jpg	Raspberry	2.0	8.0	6.0	13.0	2.0	7.0

Dimensions du jeu de données: (150, 32)

Durée de l'opération 'Création du bag of words': 8.28 s

==========
Réduction de dimension PCA
==========

Résultats de la PCA (sdf_features)
==========

root
 |-- features: vector (nullable = true)

features
[28.6589691872513...
[30.6704889986520...
[26.6684862365683...
[31.6841436746050...
[29.2510349188184...
[33.9143218745699...
[35.3151410157159...
[27.0750731114220...
[27.7020165980217...
[31.7533006032619...

Jointure entre les ids des images et les features (sdf_ima_features)
==========

```
root
 |-- image_id: long (nullable = true)
 |-- bag_of_words: vector (nullable = true)
 |-- features: vector (nullable = true)

+--------+--------------------+--------------------+
|image_id|        bag_of_words|            features|
+--------+--------------------+--------------------+
|      76|(30,[0,1,4,7,11,1...|[-2.1029243806689...|
|      60|(30,[0,1,4,7,8,14...|[-1.3391962404690...|
|      48|(30,[0,1,2,4,8,9,...|[0.77967717300582...|
|     144|(30,[0,1,2,3,4,5,...|[29.1377046696358...|
|      40|(30,[0,1,2,4,5,8,...|[1.31246907909725...|
|     110|(30,[0,1,2,3,4,5,...|[28.6589691872513...|
|      97|(30,[0,4,6,7,8,11...|[-0.7051385240676...|
|      67|(30,[0,4,7,8,11,1...|[-0.7264771357198...|
|     146|(30,[0,1,2,3,4,5,...|[29.8051167935406...|
|      19|(30,[0,1,2,3,4,5,...|[3.96508571710604...|
+--------+--------------------+--------------------+
```

Jointure entre les catégories et les features (sdf_cat_features)
==========

```
root
 |-- IMA: string (nullable = true)
 |-- image_id: long (nullable = true)
 |-- features: vector (nullable = true)
 |-- cat: string (nullable = true)

+--------------------+--------+--------------------+---------+
|                 IMA|image_id|            features|      cat|
+--------------------+--------+--------------------+---------+
|    Orange_2_100.jpg|      99|[-1.9381004478848...|   Orange|
|  Orange_102_100.jpg|      53|[-0.5822993253601...|   Orange|
|  Orange_124_100.jpg|      77|[-1.3910691745482...|   Orange|
|   Orange_24_100.jpg|      93|[-1.3747704461391...|   Orange|
|Raspberry_17_100.jpg|     108|[26.6684862365683...|Raspberry|
|     Corn_11_100.jpg|       2|[2.23326783378635...|     Corn|
|     Corn_15_100.jpg|       6|[1.16132418870859...|     Corn|
|     Corn_18_100.jpg|       9|[3.17218229130792...|     Corn|
|     Corn_44_100.jpg|      23|[2.81897891583134...|     Corn|
|  Orange_121_100.jpg|      74|[-0.8462483779417...|   Orange|
+--------------------+--------+--------------------+---------+
```

Encodage de la variable catégories (sdf_lab_features)
==========

```
root
```

```
 |-- label: double (nullable = false)
 |-- features: vector (nullable = true)

+-----+--------------------+
|label|            features|
+-----+--------------------+
|  1.0|[-1.9381004478848...|
|  1.0|[-0.5822993253601...|
|  1.0|[-1.3910691745482...|
|  1.0|[-1.3747704461391...|
|  2.0|[26.6684862365683...|
|  0.0|[2.23326783378635...|
|  0.0|[1.16132418870859...|
|  0.0|[3.17218229130792...|
|  0.0|[2.81897891583134...|
|  1.0|[-0.8462483779417...|
+-----+--------------------+


Bag of words après réduction de dimension (df_lab_features)
============================================================

   label          0         1         2         3         4         5
0    1.0  -1.938100  4.373330  7.230668 -6.354446  6.046578 -7.506347
1    1.0  -0.582299  3.418589  5.567492 -7.003518  3.878954 -5.515348
2    1.0  -1.391069  3.003477  5.123474 -6.696992  4.103136 -6.786919
3    1.0  -1.374770  3.939350  5.026782 -4.804279  1.135001 -5.934161
4    2.0  26.668486  7.078681 -0.799438 -4.180231  5.872323 -3.987865
5    0.0   2.233268 -2.347001 -2.344602 -3.261785  3.304597 -6.851918
6    0.0   1.161324 -1.819158  0.087453 -6.264102  5.091056 -5.406186
7    0.0   3.172182  0.122427  0.194158 -9.281862  8.285937 -5.758286
8    0.0   2.818979 -4.847871 -2.529770 -4.401651  4.233620 -6.630153

Dimensions du nouveau jeu de données avec les étiquettes (df_lab_features):
(150, 22)

Durée de l'opération 'Réduction de dimension': 72.5 s


==================
Classification MLP
==================

Modèle:

 MultilayerPerceptronClassificationModel:
uid=MultilayerPerceptronClassifier_572046d74e49, numLayers=4, numClasses=3,
numFeatures=21

Durée de l'opération 'Classification': 88.62 s


==================================================
> > > > TRAITEMENT DU JEU DE DONNEES DE TEST < < < <
```

```
==================================================


============================================================
Identification des chemins d'accès aux répertoires d'images
============================================================

Nombre d'images par catégorie (sous-répertoire):

    Catégorie  Nombre d'images
0       Corn               25
1  Raspberry               25
2     Orange               25

Nombre total d'images: 75

dataset_path = data/fruits_360_v3b/Test/

image_path =
data/fruits_360_v3b/Test/Corn,data/fruits_360_v3b/Test/Raspberry,data/fruits_360
_v3b/Test/Orange

Nombre de catégories de fruits: 3

2 premières catégories: ['data/fruits_360_v3b/Test/Corn',
'data/fruits_360_v3b/Test/Raspberry']
2 dernières catégories: ['data/fruits_360_v3b/Test/Raspberry',
'data/fruits_360_v3b/Test/Orange']

Durée de l'opération 'Récupération des images - Test': 0.01 s


=======================
Calcul des descripteurs
=======================


Chargement des images (rdd_images)
==================================

MapPartitionsRDD[651] at javaToPython at NativeMethodAccessorImpl.java:0

Nombre de partitions: 3
Dimension: 75

Catégories / Images / Descripteurs (rdd_cat_ima_desc)
=====================================================

PythonRDD[656] at RDD at PythonRDD.scala:53

Catégories / Images / Descripteurs (rdd_cat_ima_desc_f)
=======================================================

PythonRDD[657] at RDD at PythonRDD.scala:53
```

```
Catégories (rdd_cat)
====================

PythonRDD[661] at RDD at PythonRDD.scala:53

Identifiants des images (rdd_ima)
=================================

PythonRDD[662] at RDD at PythonRDD.scala:53

Descripteurs (rdd_desc)
=======================

PythonRDD[663] at RDD at PythonRDD.scala:53

Nombre de partitions: 3
Dimension: 5223

Collecte des catégories d'images (list_cat)
===========================================

3 premières occurences: ['Raspberry', 'Raspberry', 'Raspberry']

Collecte des identifiants des images (list_ima)
===============================================

3 premières occurences: ['Raspberry_80_100.jpg', 'Raspberry_80_100.jpg',
'Raspberry_80_100.jpg']
df_ima_cat: (5223, 2)
df_ima_cat (sans dup): (75, 2)

Identifiants des images et des catégories (sdf_ima_cat)
=======================================================

root
 |-- ima: string (nullable = true)
 |-- cat: string (nullable = true)

+--------------------+---------+
|                 ima|      cat|
+--------------------+---------+
|Raspberry_80_100.jpg|Raspberry|
|Raspberry_77_100.jpg|Raspberry|
|Raspberry_79_100.jpg|Raspberry|
+--------------------+---------+

Durée de l'opération 'Extraction des descripteurs des images - Test': 3.33 s


=======================================
Prédictions des descripteurs avec K-Means
=======================================
```

```
Prédictions (rdd_km_pred)
=========================

PythonRDD[672] at RDD at PythonRDD.scala:53

Nombre de partitions: 3
Dimension: 5223

Collecte des prédictions (list_km_pred)
=======================================

[3, 3, 3, 25, 7, 19, 11, 0, 27, 1]

Durée de l'opération 'Prédiction K-Means - Test': 5.09 s


========================
Création du bag of words
========================


Encodage des identifiants d'images et concatenation avec les prédictions
(clusters K-Means)
================================================================================
===========

Encodage des identifiants d'images (sdf_ima_label)
--------------------------------------------------

+--------------------+--------+----------+
|                 IMA|image_id|prediction|
+--------------------+--------+----------+
|Raspberry_80_100.jpg|      55|         3|
|Raspberry_80_100.jpg|      55|         3|
|Raspberry_80_100.jpg|      55|         3|
|Raspberry_80_100.jpg|      55|        25|
|Raspberry_80_100.jpg|      55|         7|
|Raspberry_80_100.jpg|      55|        19|
|Raspberry_80_100.jpg|      55|        11|
|Raspberry_80_100.jpg|      55|         0|
|Raspberry_80_100.jpg|      55|        27|
|Raspberry_80_100.jpg|      55|         1|
+--------------------+--------+----------+

Prédictions (clusters K-Means) par image (sdf_ima_pred)
-------------------------------------------------------

root
 |-- id: long (nullable = true)
 |-- prediction: long (nullable = true)

+---+----------+
| id|prediction|
```

```
+---+----------+
| 55|         3|
| 55|         3|
| 55|         3|
| 55|        25|
| 55|         7|
| 55|        19|
| 55|        11|
| 55|         0|
| 55|        27|
| 55|         1|
+---+----------+


Liste des clusters par image (Map + reduceByKey)
================================================

Clusters par image (rdd_words)
------------------------------

PythonRDD[702] at RDD at PythonRDD.scala:53

Nombre de partitions: 2
Dimension: 75

Liste de 'words' par image (sdf_worcds)
---------------------------------------

root
 |-- image_id: long (nullable = true)
 |-- words: array (nullable = true)
 |    |-- element: string (containsNull = true)

+--------+--------------------+
|image_id|               words|
+--------+--------------------+
|      54|[3, 3, 3, 25, 21,...|
|      56|[13, 3, 3, 25, 21...|
|      60|[19, 25, 7, 27, 3...|
|      66|[19, 25, 25, 7, 2...|
|      62|[19, 19, 25, 15, ...|
|      58|[19, 3, 13, 3, 19...|
|      64|[25, 25, 15, 3, 3...|
|      68|[25, 13, 7, 3, 21...|
|      52|[19, 3, 3, 3, 19,...|
|      70|[7, 3, 27, 27, 11...|
+--------+--------------------+


Création du bag of words à partir des listes de 'words' associées aux images
(CountVectorizer)
================================================================================
==============
```

```
Bag of words (sdf_bow)
----------------------

root
 |-- image_id: long (nullable = true)
 |-- bag_of_words: vector (nullable = true)

+--------+--------------------+
|image_id|        bag_of_words|
+--------+--------------------+
|      54|(30,[0,1,2,3,4,5,...|
|      56|(30,[0,1,2,3,4,5,...|
|      60|(30,[0,1,2,3,4,5,...|
|      66|(30,[0,1,2,3,4,5,...|
|      62|(30,[0,1,2,3,4,5,...|
|      58|(30,[0,1,2,3,4,5,...|
|      64|(30,[0,1,2,3,4,5,...|
|      68|(30,[0,1,2,3,4,5,...|
|      52|(30,[0,2,3,4,5,6,...|
|      70|(30,[0,1,2,3,4,5,...|
+--------+--------------------+


==========================
Sauvegarde du bag of words
==========================


Bag of words (df_bow)
=====================

   image_id                                        bag_of_words
0        54  (5.0, 5.0, 5.0, 13.0, 9.0, 9.0, 11.0, 9.0, 10....
1        56  (3.0, 8.0, 3.0, 6.0, 7.0, 10.0, 9.0, 5.0, 9.0,...
2        60  (6.0, 4.0, 3.0, 9.0, 7.0, 9.0, 8.0, 6.0, 6.0, ...
3        66  (1.0, 4.0, 5.0, 5.0, 9.0, 6.0, 3.0, 9.0, 10.0,...
4        62  (4.0, 3.0, 2.0, 9.0, 11.0, 5.0, 7.0, 10.0, 9.0...


Bag of words
============

                    ima        cat    0    1    2     3    4     5
0  Raspberry_80_100.jpg  Raspberry  1.0  8.0  3.0   8.0  9.0  14.0
1  Raspberry_77_100.jpg  Raspberry  3.0  3.0  2.0   7.0  4.0  14.0
2  Raspberry_79_100.jpg  Raspberry  5.0  5.0  5.0  13.0  9.0   9.0
3  Raspberry_81_100.jpg  Raspberry  3.0  8.0  3.0   6.0  7.0  10.0
4  Raspberry_82_100.jpg  Raspberry  5.0  7.0  5.0   7.0  9.0  13.0

Dimensions du jeu de données: (75, 32)

Durée de l'opération 'Création du bag of words - Test': 4.88 s

==========================
```

Réduction de dimension PCA
==========================

Résultats de la PCA (sdf_features)
==================================

```
root
 |-- features: vector (nullable = true)
```

```
+--------------------+
|            features|
+--------------------+
|[30.0510914065737...|
|[26.1523344783714...|
|[23.9139131399411...|
|[25.5135494776514...|
|[26.4668964038766...|
|[30.1364368364823...|
|[24.6132670646011...|
|[25.5823342822967...|
|[24.7349439639706...|
|[25.4028440202219...|
+--------------------+
```

Jointure entre les ids des images et les features (sdf_ima_features)
==

```
root
 |-- image_id: long (nullable = true)
 |-- bag_of_words: vector (nullable = true)
 |-- features: vector (nullable = true)
```

```
+--------+--------------------+--------------------+
|image_id|        bag_of_words|            features|
+--------+--------------------+--------------------+
|       6|(30,[0,1,3,4,5,6,...|[1.88671564762636...|
|      10|(30,[0,1,3,6,7,9,...|[0.39603563099539...|
|       5|(30,[0,1,2,3,4,7,...|[2.14251332966443...|
|      30|(30,[0,1,2,3,6,9,...|[-0.6024618689648...|
|      41|(30,[0,1,2,3,6,8,...|[0.67371931039150...|
|      54|(30,[0,1,2,3,4,5,...|[30.0510914065737...|
|      26|(30,[0,1,2,6,9,14...|[-1.5597208874028...|
|      68|(30,[0,1,2,3,4,5,...|[25.5823342822967...|
|      20|(30,[0,1,2,3,5,6,...|[2.50937028459530...|
|      35|(30,[0,1,2,3,5,10...|[-0.4043484642786...|
+--------+--------------------+--------------------+
```

Jointure entre les catégories et les features (sdf_cat_features)
==

```
root
```

```
 |-- IMA: string (nullable = true)
 |-- image_id: long (nullable = true)
 |-- features: vector (nullable = true)
 |-- cat: string (nullable = true)

+--------------------+--------+--------------------+---------+
|                 IMA|image_id|            features|      cat|
+--------------------+--------+--------------------+---------+
|Raspberry_87_100.jpg|      62|[26.4668964038766...|Raspberry|
|    Orange_3_100.jpg|      35|[-0.4043484642786...|   Orange|
|Raspberry_82_100.jpg|      57|[28.6641195896339...|Raspberry|
|      Corn_2_100.jpg|       9|[5.18647680516223...|     Corn|
|   Orange_43_100.jpg|      39|[-0.0956747951576...|   Orange|
|Raspberry_100_100...|      50|[24.8632082783237...|Raspberry|
|Raspberry_98_100.jpg|      73|[25.6584718118259...|Raspberry|
|     Corn_20_100.jpg|       1|[3.64587116333794...|     Corn|
|   Orange_39_100.jpg|      34|[-1.0479874383580...|   Orange|
|     Corn_29_100.jpg|       8|[2.32739838521921...|     Corn|
+--------------------+--------+--------------------+---------+


Encodage de la variable catégories (sdf_lab_features)
=====================================================

root
 |-- label: double (nullable = false)
 |-- features: vector (nullable = true)

+-----+--------------------+
|label|            features|
+-----+--------------------+
|  2.0|[26.4668964038766...|
|  1.0|[-0.4043484642786...|
|  2.0|[28.6641195896339...|
|  0.0|[5.18647680516223...|
|  1.0|[-0.0956747951576...|
|  2.0|[24.8632082783237...|
|  2.0|[25.6584718118259...|
|  0.0|[3.64587116333794...|
|  1.0|[-1.0479874383580...|
|  0.0|[2.32739838521921...|
+-----+--------------------+


Bag of words après réduction de dimension (df_lab_features)
===========================================================

   label          0         1          2         3         4         5
0    2.0  26.466896 -0.037070   4.952361 -8.969748 -4.504904 -2.318564
1    1.0  -0.404348  4.709597   7.649410 -2.661897 -1.153152 -2.915067
2    2.0  28.664120  0.477897  13.081357 -1.182491 -0.783142 -5.041057
3    0.0   5.186477 -2.816440   3.138431 -2.043061 -1.124801 -3.906776
4    1.0  -0.095675  6.139206   8.612166 -3.342973 -2.877894 -3.623146
5    2.0  24.863208  3.969479   0.115875 -5.347710 -0.081041 -2.178174
```

```
6    2.0  25.658472  4.911228   0.661215 -3.971885  5.415495 -5.078884
7    0.0   3.645871 -3.190890   4.703170 -1.548317 -3.569367 -6.738843
8    1.0  -1.047987  4.515285   6.073193 -3.585146  0.104125 -2.443161

Dimensions du nouveau jeu de données avec les étiquettes (df_lab_features): (75,
22)

Durée de l'opération 'Réduction de dimension - Test': 50.54 s


===============
Prédictions MLP
===============


Prédictions (test_lab_pred)
===========================

DataFrame[features: vector, prediction: double]
+--------------------+----------+
|            features|prediction|
+--------------------+----------+
|[30.0510914065737...|       2.0|
|[26.1523344783714...|       2.0|
|[23.9139131399411...|       2.0|
|[25.5135494776514...|       2.0|
|[26.4668964038766...|       2.0|
|[30.1364368364823...|       2.0|
|[24.6132670646011...|       2.0|
|[25.5823342822967...|       2.0|
|[24.7349439639706...|       2.0|
|[25.4028440202219...|       2.0|
+--------------------+----------+


Prédictions (predictionAndLabels_2)
===================================

DataFrame[prediction: double, label: double, features: vector]
+----------+-----+--------------------+
|prediction|label|            features|
+----------+-----+--------------------+
|       2.0|  2.0|[26.4668964038766...|
|       1.0|  1.0|[-0.4043484642786...|
|       2.0|  2.0|[28.6641195896339...|
|       2.0|  0.0|[5.18647680516223...|
|       1.0|  1.0|[-0.0956747951576...|
|       2.0|  2.0|[24.8632082783237...|
|       2.0|  2.0|[25.6584718118259...|
|       2.0|  0.0|[3.64587116333794...|
|       1.0|  1.0|[-1.0479874383580...|
|       1.0|  0.0|[2.32739838521921...|
+----------+-----+--------------------+
```

```
Test set accuracy (MLP) = 0.72

Durée de l'opération 'Prédiction - Test': 18.71 s


==========
Evaluation
==========


Jointure entre les identifiants des images et les features (sdf_ima_features)
================================================================================

root
 |-- image_id: long (nullable = true)
 |-- bag_of_words: vector (nullable = true)
 |-- features: vector (nullable = true)
 |-- prediction: double (nullable = false)

+--------+--------------------+--------------------+----------+
|image_id|        bag_of_words|            features|prediction|
+--------+--------------------+--------------------+----------+
|       6|(30,[0,1,3,4,5,6,...|[1.88671564762636...|       0.0|
|      10|(30,[0,1,3,6,7,9,...|[0.39603563099539...|       1.0|
|       5|(30,[0,1,2,3,4,7,...|[2.14251332966443...|       1.0|
|      30|(30,[0,1,2,3,6,9,...|[-0.6024618689648...|       1.0|
|      41|(30,[0,1,2,3,6,8,...|[0.67371931039150...|       1.0|
|      54|(30,[0,1,2,3,4,5,...|[30.0510914065737...|       2.0|
|      26|(30,[0,1,2,6,9,14...|[-1.5597208874028...|       1.0|
|      68|(30,[0,1,2,3,4,5,...|[25.5823342822967...|       2.0|
|      20|(30,[0,1,2,3,5,6,...|[2.50937028459530...|       2.0|
|      35|(30,[0,1,2,3,5,10...|[-0.4043484642786...|       1.0|
+--------+--------------------+--------------------+----------+


Jointure entre les catégories et les features (sdf_cat_features)
================================================================

root
 |-- IMA: string (nullable = true)
 |-- image_id: long (nullable = true)
 |-- features: vector (nullable = true)
 |-- prediction: double (nullable = false)
 |-- cat: string (nullable = true)

+--------------------+--------+--------------------+----------+---------+
|                 IMA|image_id|            features|prediction|      cat|
+--------------------+--------+--------------------+----------+---------+
|Raspberry_87_100.jpg|      62|[26.4668964038766...|       2.0|Raspberry|
|    Orange_3_100.jpg|      35|[-0.4043484642786...|       1.0|   Orange|
|Raspberry_82_100.jpg|      57|[28.6641195896339...|       2.0|Raspberry|
|      Corn_2_100.jpg|       9|[5.18647680516223...|       2.0|     Corn|
|   Orange_43_100.jpg|      39|[-0.0956747951576...|       1.0|   Orange|
|Raspberry_100_100...|      50|[24.8632082783237...|       2.0|Raspberry|
```

```
|Raspberry_98_100.jpg|      73|[25.6584718118259...|       2.0|Raspberry|
|     Corn_20_100.jpg|       1|[3.64587116333794...|       2.0|     Corn|
|   Orange_39_100.jpg|      34|[-1.0479874383580...|       1.0|   Orange|
|     Corn_29_100.jpg|       8|[2.32739838521921...|       1.0|     Corn|
+--------------------+--------+--------------------+----------+---------+


Encodage de la variable catégories (sdf_lab_features)
=====================================================

root
 |-- ima: string (nullable = true)
 |-- cat: string (nullable = true)
 |-- label: double (nullable = false)
 |-- prediction: double (nullable = false)
 |-- features: vector (nullable = true)

+--------------------+---------+-----+----------+--------------------+
|                 ima|      cat|label|prediction|            features|
+--------------------+---------+-----+----------+--------------------+
|Raspberry_87_100.jpg|Raspberry|  2.0|       2.0|[26.4668964038766...|
|    Orange_3_100.jpg|   Orange|  1.0|       1.0|[-0.4043484642786...|
|Raspberry_82_100.jpg|Raspberry|  2.0|       2.0|[28.6641195896339...|
|      Corn_2_100.jpg|     Corn|  0.0|       2.0|[5.18647680516223...|
|   Orange_43_100.jpg|   Orange|  1.0|       1.0|[-0.0956747951576...|
|Raspberry_100_100...|Raspberry|  2.0|       2.0|[24.8632082783237...|
|Raspberry_98_100.jpg|Raspberry|  2.0|       2.0|[25.6584718118259...|
|     Corn_20_100.jpg|     Corn|  0.0|       2.0|[3.64587116333794...|
|   Orange_39_100.jpg|   Orange|  1.0|       1.0|[-1.0479874383580...|
|     Corn_29_100.jpg|     Corn|  0.0|       1.0|[2.32739838521921...|
+--------------------+---------+-----+----------+--------------------+


Catégories réelles (label) vs Prédictions (prediction)
======================================================

prediction   0.0    1.0    2.0
label
0.0         16.0   56.0   28.0
1.0          0.0  100.0    0.0
2.0          0.0    0.0  100.0

Durée de l'opération 'Evaluation - Test': 50.56 s

Durée de l'opération 'Fin des traitements': 0.0 s
Durée totale de traitement: 00 h 06 m 00 s

Durée des opérations
--------------------
                                      Opération  Durée Estimation
0                         Récupération des images   0.01
1          Extraction des descripteurs des images  18.66
2                              Clustering K-Means  32.76
3                              Prédiction K-Means   6.20
```

```
4                         Création du bag of words   8.28
5                           Réduction de dimension  72.50
6                                    Classification  88.62
7                    Récupération des images - Test   0.01
8   Extraction des descripteurs des images - Test   3.33
9                         Prédiction K-Means - Test   5.09
10                Création du bag of words - Test   4.88
11                  Réduction de dimension - Test  50.54
12                               Prédiction - Test  18.71
13                               Evaluation - Test  50.56
14                             Fin des traitements   0.00

====================================
> > > > Traitements finalisés < < < <
====================================
```