

# Projet Classification Fruit



## I- Objectif :

Effectuer une classification de différents types de fruits. La machine reçoit en input des images (samples) et doit ressortir en output les noms des fruits (labels). Notre objectif est d'avoir une accuracy de 90% minimum.

## II- Dataset :

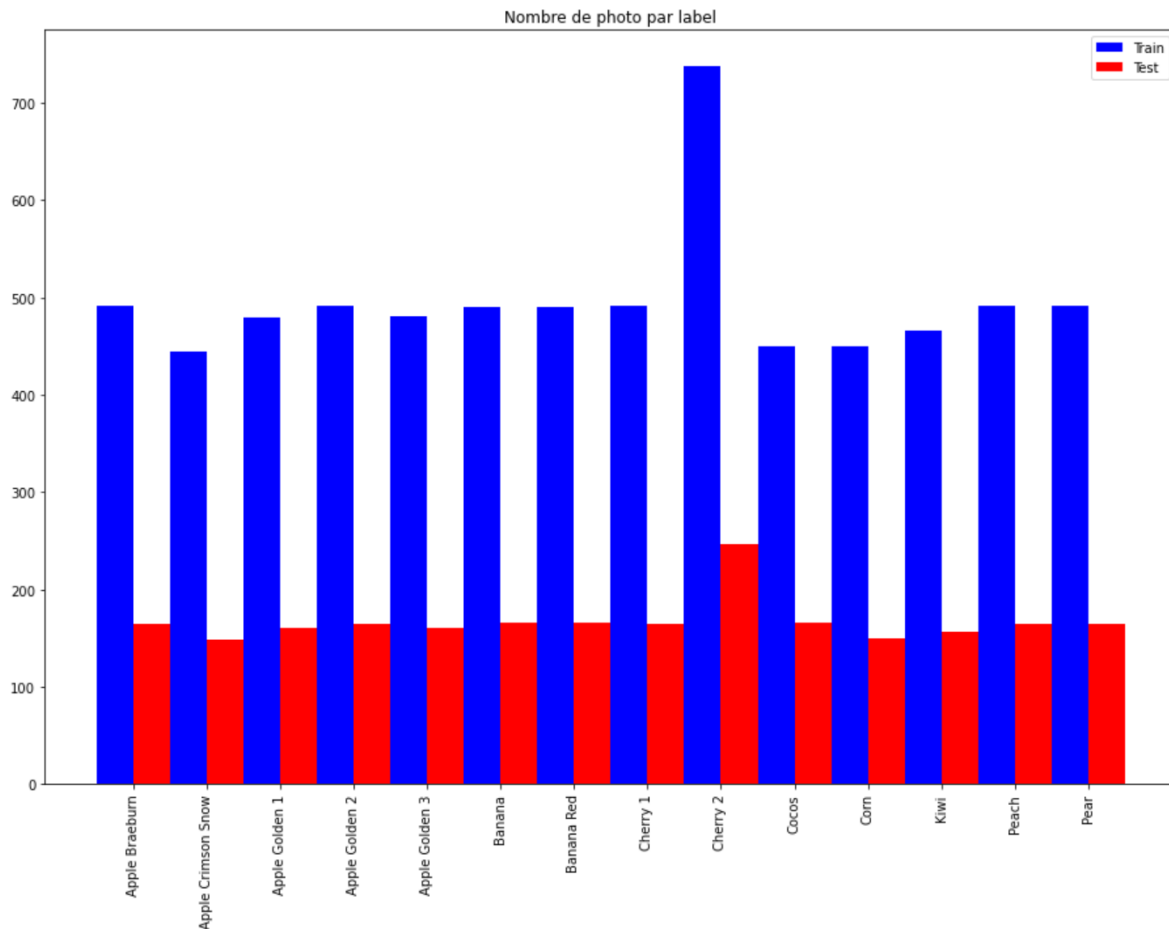
Nous avons choisi le dataset "Fruits 360" qui contient diverses images de différents types de fruits, pris sous des angles différents et qui provient du site Kaggle (<https://www.kaggle.com/moltean/fruits> ).

L'ensemble des images ont été vérifiées et notre Dataset ne contient aucune erreur dans les images. Toutes les données utilisées sont pertinentes, plus nous en avons plus nous serons précis. Nous avons sélectionné quatorze fruits différents avec des variations possibles, par exemple, nous avons plusieurs "golden apple" mais ayant une apparence différente.

## 1- Taille du dataset :

Nom	Test	Train	Total
Apple Braeburn	164	492	656
Apple Crimson Snow	148	444	592
Apple Golden 1	160	480	640
Apple Golden 2	164	492	656
Apple Golden 3	161	481	642
Banana	166	490	656
Banana Red	166	490	656
Cherry 1	164	492	656
Cherry 2	246	738	984
Cocos	166	450	616
Kiwi	156	466	622
Corn	150	450	600
Peach	164	492	656
Pear	164	492	656
Total	2339	6949	9288

Sur le graphique ci-dessous, on observe le nombre d'images contenue dans le test et l'entraînement en fonction du label.



### III- SVM

Nous avons décidé d'utiliser la méthode SVM. Les algorithmes de SVM appartiennent à une famille d'algorithmes qui font appel à l'apprentissage dit supervisé, c'est -à-dire que nous lui fournissons un dataset que nous pensons pertinents avec des labels à contrario de l'apprentissage non supervisé, et c'est grâce à ces labels qu'il peut déterminer son accuracy.

Ils peuvent être adaptés à des problèmes de classification portant sur plus de 2 classes, et à des problèmes de régression linéaire. Là où un réseau de neurones demande un travail en amont pour déterminer la bonne structure et les bons paramètres à utiliser, les SVM obtiennent de bons résultats même sans préparation.

## I- Choix des paramètres :

Pour le choix des paramètres nous avons utilisé la fonction GridSearchCV de sklearn. Elle permet de parcourir les hyperparamètres prédéfinis et d'adapter l'estimateur, c'est-à-dire le modèle donc notre cas c'est le modèle SVM, à votre ensemble d'entraînement. Ainsi, au final, nous pouvons sélectionner les meilleurs paramètres parmi les hyperparamètres répertoriés.

Nous lui avons donné trois hyperparamètres C, kernel et gamma. Pour C nous lui avons donné sept valeurs (0.001, 0.01, 0.1, 1, 10, 100, 1000), pour kernel nous lui avons donné une seule valeur ('linear') et pour gamma trois valeurs (0.01, 0.001, 0.0001).

Après plusieurs essais, nous avons trouvé que les meilleurs paramètres sont de gamma = 0.01, C=1 et kernel='linear'.

### 1) Résultat avec les images en nuance de gris :

Paramètres SVM utilisés : gamma = 0.01, C=1 et kernel='linear'.

Nombre de fruit	Nom des fruits	Certitude (en %)
2	Banane, Kiwi	100
3	Banane, Kiwi, Pomme Braeburn	92
5	Banane, Kiwi, Pomme Braeburn, Pêche, Poire	92
6	Banane, Kiwi, Pomme Braeburn, Pêche, Poire, Pomme Crimson Snow	94
7	Banane, Kiwi, Pomme Braeburn, Pêche, Poire, Pomme Crimson Snow, Banane rouge	89
10	Banane, Kiwi, Pomme Braeburn, Pêche, Poire, Pomme Crimson Snow, Banane rouge, Cerise 1, Noix de coco, Maïs	87
11	Banane, Kiwi, Pomme Braeburn, Pêche, Poire, Pomme Crimson Snow, Banane rouge, Cerise 1, Noix de coco, Maïs, Pomme Golden 1	86
13	Banane, Kiwi, Pomme Braeburn, Pêche, Poire, Pomme Crimson Snow, Banane rouge, Cerise, Noix de coco, Maïs, Pomme Golden 1, 2 et 3	85
14	Banane, Kiwi, Pomme Braeburn, Pêche, Poire, Pomme Crimson Snow, Banane rouge, Cerise 1 et 2, Noix de coco, Maïs, Pomme Golden 1, 2 et 3	88

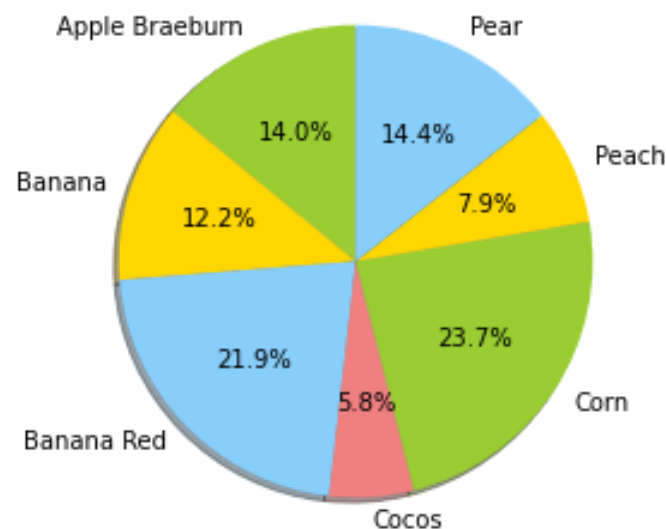
Nous avons décidé d'ajouter des fruits graduellement, comme attendu, le pourcentage diminue lorsque que l'on ajoute des fruits, mais lorsqu'un autre type de pomme a été ajouté, on observe que le pourcentage de certitude a augmenté.

Quand nous avons ajouté un autre type de banane, le pourcentage de certitude a diminué de 5% cela est dû aux nombreuses erreurs faites sur ce fruit qui a un aspect identique à son homologue.

Quand nous avons ajouté le maïs (corn), le pourcentage de certitude a diminué de 2% cela est dû aux nombreuses erreurs faites sur ce fruit, qui vu du dessus ressemble énormément à une pomme.

Lorsque nous avons ajouté Cerise 2, le pourcentage de certitude à augmenté de 3%, cela est peut-être dû au fait que la machine avait un autre type de cerise pour son entraînement et donc les erreurs faites avec la Cerise 1 ont complètement diminué.

**Erreur par label sur la totalité des erreurs :**



## 2) Résultat avec les images en couleur (RGB) :

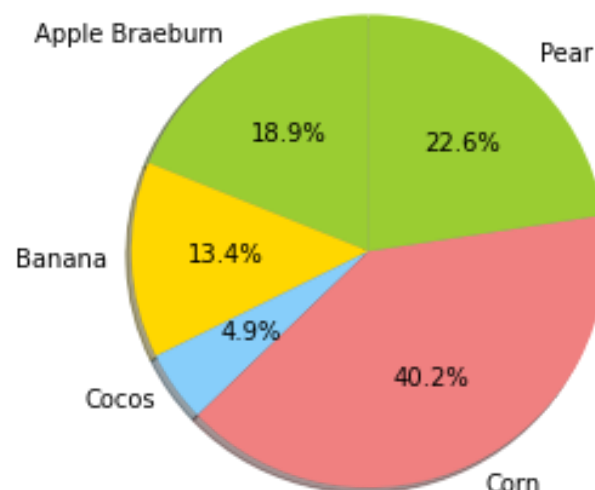
Paramètres SVM utilisés :  $\gamma = 0.01$ ,  $C=1$  et  $\text{kernel}='linear'$ .

Nombre de fruit	Nom des fruits	Certitude (en %)
14	Banane, Kiwi, Pomme Braeburn, Pêche, Poire, Pomme Crimson Snow, Banane rouge, Cerise 1 et 2, Noix de coco, Maïs, Pomme Golden 1, 2 et 3	92.9

Sans surprise, lorsque les images sont en couleur, le degré de certitude augmente. Néanmoins, le temps d'exécution est bien plus élevé.

Quand nous sommes passés en couleur, le nombre de labels contenant des erreurs à diminué de sept à cinq. Le maïs (corn) est passé de 23,7% à 40,2% ce qui peut être dû à la couleur et à la forme proche de celle des pommes et d'autres fruits. A contrario, il n'y a eu aucune erreur avec Banana Red qui avait été confondu avec la Banana, cela est sûrement dû au fait qu'il n'y a que ce fruit qui présente les caractéristiques suivantes : couleur rouge avec une forme en croissant de lune.

**Erreur par label sur la totalité des erreurs :**



## IV- Conclusion

Pour conclure, en utilisant la méthode SVM, sur un dataset comprenant 14 fruits et avec 9288 images en couleurs, nous avons atteint notre objectif avec une précision de 92.9%.