

# Projet Vic-2-I

*(Vegetables In Cabinets To (alimentary) Independence, a.k.a Victuailles)*

## avec PopFarm

Projet en équipe de 2 à 3 personnes.

### 1. Avant-propos

Dans le cadre du programme Agritech entre Hectar et 42, le projet Vic-2-I est effectué en collaboration avec [PopFarm](#).

### 2. Contexte agricole : les fermes urbaines

Avec la tendance grandissante des circuits courts et des productions bio et écoresponsable, [PopFarm](#) a imaginé une machine automatique permettant la production locale de certains fruits et légumes suivant ces principes. Basilic, fraises, ciboulette, et tant d'autres sont ainsi cultivés de façon automatique par des enseignes alimentaires qui les mettent directement en rayon. La machine, bardée de capteurs, est dotée d'une IA qui s'assure du bon développement des plantes.

Un des principaux défis dans la conception d'IA dans le monde réel est, à partir d'une base de données annotées, être capable de bien les traiter pour entraîner différents types de modèles pour qualifier cette donnée. Un élément essentiel est donc de pouvoir bien mesurer les performances des modèles, et pour cela, une des meilleures solutions est de les confronter à des experts.

Au sein de cet exercice, nous vous proposons de travailler sur les données qui nous ont servi à élaborer un POC d'une application de reconnaissance de plantes et de détection de maladie.

Ce projet s'axe sur les trois parties essentielles à la mise en place d'une IA : traitement de la donnée, modélisation, et visualisation/interprétation de ces deux premières.

Le fil directeur de chaque partie est d'appliquer des méthodes d'analyse des données avec pour objectif de pouvoir faire un retour clair à nos équipes d'agronomes, non seulement pour les assister, mais aussi pour obtenir leurs observations qui permettront de détecter des erreurs ou d'améliorer la solution en continu.

### 3. Les données

Les données utilisées dans ce projet sont celles utilisées dans le papier "J, ARUN PANDIAN; GOPAL, GEETHARAMANI (2019), "Data for: Identification of Plant Leaf Diseases Using a 9-layer Deep Convolutional Neural Network", Mendeley Data, V1, doi: 10.17632/tywbtsjrjv.1".

Dans ce jeu de données, 39 classes différentes d'images de feuilles et d'arrière-plan de plantes sont disponibles. L'ensemble de données contenant 61 486 images. Les auteurs ont

augmenté artificiellement ces données avec six techniques. Les techniques sont le retournement d'image, la correction gamma, l'injection de bruit, l'augmentation de couleur PCA, la rotation et la mise à l'échelle.

Les différentes classes sont,

- 1.Apple\_scab
- 2.Apple\_black\_rot
- 3.Apple\_cedar\_apple\_rust
- 4.Apple\_healthy
- 5.Background\_without\_leaves
- 6.Blueberry\_healthy
- 7.Cherry\_powdery\_mildew
- 8.Cherry\_healthy
- 9.Corn\_gray\_leaf\_spot
- 10.Corn\_common\_rust
- 11.Corn\_northern\_leaf\_blight
- 12.Corn\_healthy
- 13.Grape\_black\_rot
- 14.Grape\_black\_measles
- 15.Grape\_leaf\_blight
- 16.Grape\_healthy
- 17.Orange\_haunglongbing
- 18.Peach\_bacterial\_spot
- 19.Peach\_healthy
- 20.Pepper\_bacterial\_spot
- 21.Pepper\_healthy
- 22.Potato\_early\_blight
- 23.Potato\_healthy
- 24.Potato\_late\_blight
- 25.Raspberry\_healthy
- 26.Soybean\_healthy
- 27.Squash\_powdery\_mildew
- 28.Strawberry\_healthy
- 29.Strawberry\_leaf\_scorch
- 30.Tomato\_bacterial\_spot
- 31.Tomato\_early\_blight
- 32.Tomato\_healthy
- 33.Tomato\_late\_blight
- 34.Tomato\_leaf\_mold
- 35.Tomato\_septoria\_leaf\_spot
- 36.Tomato\_spider\_mites\_two-spotted\_spider\_mite
- 37.Tomato\_target\_spot
- 38.Tomato\_mosaic\_virus
- 39.Tomato\_yellow\_leaf\_curl\_virus

Le jeu de données original ainsi que l'augmenté [sont disponibles ici](#). Nous vous recommandons d'utiliser l'un et l'autre pour observer les effets de ces augmentations tout au long de l'exercice.

Les photos de chaque feuille sont dans un sous répertoire dont le nom constitue le label correspondant.

## 4. Partie 1 : traitement de la donnée, analyse et visualisation

Etudier les données initiales est primordial pour mettre en place un pipeline de machine learning efficace et fiable.

Dans cette partie, nous vous demandons donc de mettre en place des outils d'analyse de la base de données fournies. Il s'agit en autre :

- Analyser le jeu de données global: distribution des différentes classes, maladie commune à plusieurs espèces, ...
- Mettre en place différentes méthodes d'extraction directes des caractéristiques des plantes en utilisant [plantCV](#)
- Faire un reporting des différentes informations trouvées lors des étapes précédentes, graphique de distributions globales, matrices de corrélations de features avec les labels, ...
- Faire le rendu sous la forme d'une application web de type FastApi/GradIO qui permettra à nos équipes d'agronomes d'étudier et analyser les résultats obtenus, ainsi que de faire des feedbacks sous forme de prise de note

## 5. Partie 2 : classification des features

En utilisant uniquement les features obtenues lors de l'utilisation de plantCV, mettre en place différents modèles de classification des plantes et de détection de maladie. Pour cela, nous vous conseillons d'utiliser les solutions proposées par la librairie [scikit-learn](#) mais tout autre solutions peut être envisagée.

Chaque étape du pipeline de classification devra être expliquée et justifiée. Le but ici est de comparer les performances d'au moins cinq modèles différents.

Un rapport d'étude devra indiquer donc les traitements effectués, les choix de modèles, les métriques d'évaluations devront être justifiées, et, si possible, les décisions des modèles devront être interprétées.

Comme pour la première partie, le rendu final sera sous la forme d'une application web de type FastApi/GradIO qui permettra à nos équipes d'agronomes d'étudier et analyser les résultats obtenus, ainsi que de faire des feedbacks sous forme de prise de note

## 6. Partie 3 : classification avec du Deep Learning

L'un des avantages principaux du deep learning est de permettre de ne pas faire la phase d'extraction de features manuellement, le réseau pouvant s'en charger.

Dans cette partie, nous vous demandons de mettre en place différents modèles d'apprentissage profond. Pour cela, nous vous conseillons d'utiliser les solutions [Keras](#) ou [pytorch-lightning](#).

Ici encore, nous vous demandons d'explicitier et justifier le plus possible le pipeline mise en place. Nous vous demandons de prêter attention au prétraitement des jeux de données, surtout en cas de déséquilibres des classes d'apprentissage, de justifier la recherche d'hyperparamètres, les métriques utilisées.

L'une des difficultés de ces modèles est qu'ils sont considérés comme "black-box" i.e. leurs décisions ne peuvent pas être comprise directement. Des solutions existent comme Shap, Lime et d'autres. Dans la mesure du possible et si le temps le permet, ajouter ces mécanismes d'interprétation au rapport final.

Comme pour la première partie, le rendu final sera sous la forme d'une application web de type FastApi/GradIO qui permettra à nos équipes d'agronomes d'étudier et analyser les résultats obtenus, ainsi que de faire des feedbacks sous forme de prise de note

## 7. Considérations techniques et légales

L'accès aux données se fait de la façon suivante :

[\[Lien vers les données\]](#)

Ces données vous sont fournies dans le cadre du programme Agritech que vous suivez. Elles ne peuvent en aucun cas être diffusées, utilisées pour d'autres finalités que ce projet. Tout travail effectué sur ces données implique votre acceptation sans restriction des ces règles.

L'accès à la machine pour lecture live et pilotage se fait comme suit :

[...]

## 8. Suivi

L'équipe de PopFarm effectue une présentation générale du projet, puis à intervalles réguliers se propose d'interagir avec les groupes d'étudiants travaillant sur le projet. Pour cela, un channel dédié sur discord sera mis en place pour faciliter les échanges.

## 9. Rendu et évaluation

Chaque partie fait l'objet d'un rendu dans un répertoire séparé, sur le repo git attribué lors de l'inscription de l'équipe.

L'évaluation du projet sera conjointement effectuée par les équipes de PopFarm, au moins un exploitant (qui peut être un étudiant Hectar utilisateur de la machine), et un pair dans la promo Agritech.