المندوبية السامية للتخطيط
HAUT-COMMISSARIAT AU PLAN

مدرسة علوم المعلومات
ⵜⴰⵙⴷⴰⵡⵉⵜ ⵜⵓⵙⵙⵏⵉⵡⵉⵏ ⵏ ⵜⵣⵔⴼⵜ
ECOLE DES SCIENCES DE L'INFORMATION

جامعة محمد الخامس بالرباط
المعهد العلمي
Université Mohammed V de Rabat
Institut Scientifique

Rapport de Stage d'études réalisé au sein de
Institut Scientifique de Rabat

Le créneau du stage :

Artificial Intelligence for Production Optimization of Smart Agriculture

---

Préparé par : TERRAF ELMEHDI
2ème ACI (Groupe A)

Encadré par : M. Moujahdi Chouaib

Dédicace

À Dieu le tout puissant

À mes parents,

Merci pour votre soutien et votre confiance à un moment où je doutai en mes compétences. Merci pour vos sacrifices et pour votre amour inconditionnel. J'espère que je vous rendrai fiers de moi un jour, je vous suis reconnaissant à tout jamais.

À mes amis,

Un très grand merci pour votre soutien durant mes moments les plus sombres. Je vous aime du plus profond de mon cœur et vous souhaite tout le bonheur et toute la réussite qui puissent exister dans ce monde.

Veuillez trouver dans ce modeste travail l'expression de mon affection.

REMERCIEMENTS

Au terme travail, je garde avant tout une pensée à ceux qui m'ont soutenu et accompagné tout au long de ce projet et je tiens à les remercier…

Je tiens à présenter mes sincères remerciements à mon tuteur, Monsieur Moujahdi Chouaib, pour sa confiance et les connaissances qu'il a partagé avec moi. Je le remercie aussi pour sa disponibilité et la qualité de son encadrement, ainsi Mon binôme Amal EL AMRANI, nous avons formé une bonne équipe et je suis impatient de travailler à nouveau avec elle.

Un grand merci à Madame Maryem Rhanoui et Madame Mounia Mikram, pour le soutien durant toute cette période de stage.

Je remercie également toute l'équipe pédagogique de l'Ecole des Sciences de l'Information.

Que tous ceux qui ont contribué à mener à bien ce stage trouvent ici l'expression de ma parfaite considération.

Résumé :

Entre 1960 et 2015, la production agricole a triplé et la population mondiale est passée de 3 milliards à 7 milliards. Alors que la technologie a joué un rôle dans l'augmentation de la production agricole sous forme de pesticides, d'engrais et de machines, nous pouvons dire que la plupart des gains peuvent être attribués au simple fait de cultiver plus de terres, de défricher les forêts et de détourner l'eau douce vers les champs. Il va donc falloir désormais être plus débrouillard pour améliorer la production, mais cette fois de manière intelligente et durable.

L'agriculture nourrira environ 9,7 milliards de personnes dans le monde d'ici 2050, il est donc nécessaire d'adopter la transformation numérique de l'agriculture, y compris l'utilisation de l'intelligence artificielle, et d'adopter la bonne approche pour répondre à la demande croissante

L'objectif principal de ce stage est d'aborder en générale le sujet de l'utilisation des techniques classiques de Machine Learning et celles de Deep Learning pour l'agriculture numérique en mettant l'accent sur l'optimisation et l'amélioration de la production agricole.

Le présent rapport synthétise le travail effectué, qui a pour but la construction d'un modèle prédictif qui permettant de suggérer les plantations les plus appropriées en fonction des conditions climatiques et des propriétés du sol.

TABLE DE MATIÈRE

Liste des figures

Introduction

L'agriculture intelligente fait référence à la vaste application de l'intelligence artificielle (IA), qui implique le big data, l'internet des objets (IoT), l'apprentissage profond et bien d'autres technologies numériques. Avec la croissance de la population mondiale, une augmentation significative de la production alimentaire doit être réalisée. Assurer la disponibilité et la qualité constantes et cohérentes des aliments à l'échelle mondiale sans affecter les écosystèmes naturels est un défi pour les technologies modernes.

L'apprentissage profond est une nouvelle technologie de pointe pour le traitement des images et l'analyse des données. Elle a donné des résultats prometteurs, possède un énorme potentiel et a été employée avec succès dans divers domaines, dont l'agriculture.

Ces dernières années, les applications agricoles basées sur l'apprentissage profond (agriculture intelligente) ont connu un succès considérable ; il s'agit de gérer différentes activités agricoles à l'aide de données acquises à partir de diverses sources. Divers systèmes intelligents basés sur l'IA se distinguent par leur capacité à enregistrer et interpréter les données et à aider les agriculteurs à prendre les bonnes décisions au bon moment. Les données peuvent être enregistrées à l'aide de nœuds IoT installés (capteurs), traitées par n'importe quelle méthode de deep learning et imposer des décisions sur les zones opérationnelles par le biais d'actionneurs.

Avant d'entamer la réalisation de ce projet, nous arrêterons sur les tendances, les problèmes et les orientations futures de la prédiction des plantations à l'ère de deep learning, puis nous allons nous intéresser à la documentation des différents algorithmes, à savoir, les algorithmes d'apprentissage supervisé et non supervisé et les algorithmes d'apprentissages profond.

L'ORGANISME DU STAGE

Présentation de l'institut scientifique de Rabat

L'Institut scientifique de Rabat est un établissement universitaire de *Recherche scientifique*, appartenant à l'Université Mohammed V de Rabat et sous tutelle du ministère de l'Enseignement Supérieur, de la Recherche Scientifique et de la Formation des Cadres. Fondé il y a presque un siècle, il est l'institution scientifique la plus ancienne du pays.

Officiellement chargé d'effectuer dans le domaine des sciences de la nature des recherches fondamentales, en rapport avec le sol, la faune et la flore. Il est aussi chargé de dresser l'inventaire du milieu biologique et physique, de constituer des collections du Musée National, d'aménager des laboratoires, et les stations nécessaires à ses recherches. La constitution d'une bibliothèque scientifique fait également partie de ses principaux objectifs.

Figure 1: Organigramme de l'ISR

L'Institut Scientifique est un éditeur scientifique, activité à laquelle il doit beaucoup son rayonnement à l'échelle nationale et internationale. Ses publications paraissent depuis sa création en 1920 jusqu'à nos jours ; actuellement elles sont au nombre de sept :

- Bulletin de l'Institut Scientifique (Sciences de la Terre), annuel et indexé dans GoogleScholar, Georef (USA, spécial géologie), DOAJ (Suède), Zoological Record (Thomson-Reuters) et Scopus.
- Bulletin de l'Institut Scientifique (Sciences de la Vie), annuel et indexé dans GoogleScholar, DOAJ et Zoological Record.
- Travaux de l'Institut Scientifique (série générale), parution irrégulière.
- Travaux de l'Institut Scientifique (série géologie et géographie physique), parution irrégulière.
- Travaux de l'Institut Scientifique (série zoologie), parution irrégulière.
- Travaux de l'Institut Scientifique (série botanique), parution irrégulière.
- Documents de l'Institut Scientifique, parution irrégulière.

Les activités développées à l'ISR visent à faire l'inventaire de l'ensemble des ressources naturelles du Maroc (faune, flore et ressources du sous-sol), suivant les priorités nationales en matière de conservation du patrimoine naturel, de l'environnement, de l'éducation et du développement.

PATRIMOINE BOTANIQUE DE LA BIBLIOTHÈQUE DE L'ISR	
Types	Nombre (approximatif)
Ouvrages	2500
Collections de périodiques	368
Spécimens de l'herbier	160 000
Plantes	1200

Figure 2:Patrimoine de la bibliothèque de botanique et d'écologie végétale de l'ISR

Partie théorique :

Problématique

Le problème commun des agriculteurs est qu'ils ne choisissent pas la bonne semence en fonction des exigences du sol. De ce fait, ils subissent un sérieux revers en termes de productivité. L'agriculture de précision est une technique agricole moderne qui utilise des données de recherche sur les caractéristiques et les types de sol, ainsi que des données sur le rendement de la récolte, et propose aux agriculteurs la bonne semence en fonction des paramètres spécifiques de leur sol. Cela réduit le mauvais choix d'une plante et augmente la productivité. Ce projet, a pour but d'aider les agriculteurs à prendre une décision éclairée sur la récolte à faire en fonction de la saison des semailles, de la situation géographique de l'exploitation et des caractéristiques du sol.

La décision d'un agriculteur quant au choix de la culture est généralement influencée par son intuition et d'autres facteurs non pertinents tels que la recherche de profits immédiats, le manque de connaissance de la demande du marché, la surestimation du potentiel d'un sol à supporter une culture particulière, etc. Une décision très malencontreuse de la part de l'agriculteur peut entraîner des répercussions importantes sur la situation financière de sa famille, mais aussi sur l'ensemble de l'économie du pays. Pour cette raison, nous avons identifié le dilemme de l'agriculteur sur le choix de la culture à faire pendant une saison particulière comme un problème très grave. Fournir des informations prédictives aux agriculteurs, les aidant ainsi à prendre une décision éclairée sur le choix de la culture. Prendre en compte les paramètres environnementaux (température, précipitations) et les caractéristiques du sol (pH, phosphore, potassium,) avant de recommander la culture la plus adaptée à l'utilisateur.

Nous avons besoin de construire un modèle prédictif qui permettant de suggérer les plantations les plus appropriées en fonction des conditions climatiques et des propriétés du sol.

Le deep learning pour l’agriculture :

Le Machine Learning est un sous-domaine de l'IA qui permet au système d'apprendre à partir de concepts et de connaissances sans être explicitement programmé. L'apprentissage automatique est une application de l'intelligence artificielle (IA) qui donne aux systèmes la capacité d'apprendre automatiquement et d'évoluer à partir de l'expérience sans être spécialement programmés par le programmeur. Le processus d'apprentissage commence par des observations ou des données, telles que des exemples, une expérience directe ou des instructions, afin de rechercher des modèles dans les données et de prendre de meilleures décisions à l'avenir sur la base des exemples que nous fournissons. L'objectif principal de l'apprentissage automatique est de permettre aux ordinateurs

d'apprendre automatiquement et d'ajuster leurs actions pour améliorer la précision et l'utilité du programme, sans aucune intervention ou assistance humaine.

Les techniques de machine learning peuvent être classées dans les catégories suivantes :

L'apprentissage supervisé prend un ensemble de paires caractéristiques, appelé ensemble d'apprentissage (training set). À partir de cet ensemble d'apprentissage, le système crée un modèle généralisé de la relation entre l'ensemble des caractéristiques descriptives et les caractéristiques cibles sous la forme d'un programme contenant un ensemble de règles. L'objectif est d'utiliser le programme de sortie produit pour prédire l'étiquette d'un ensemble de caractéristiques d'entrée non étiquetées et non vues auparavant, c'est-à-dire pour prédire le résultat de nouvelles données.

L'apprentissage non supervisé prend un ensemble de données de caractéristiques descriptives sans étiquettes comme ensemble d'entraînement. Dans l'apprentissage non supervisé, les algorithmes sont laissés à eux-mêmes pour découvrir des structures intéressantes dans les données. L'objectif est maintenant de créer un modèle qui trouve une structure cachée dans l'ensemble de données, comme des clusters ou des associations naturelles.

Le machine learning semi-supervisé se situe quelque part entre l'apprentissage supervisé et l'apprentissage non supervisé, car il utilise à la fois des données étiquetées et non étiquetées pour la formation, généralement une petite quantité de données étiquetées et une grande quantité de données non étiquetées. Les systèmes qui utilisent cette méthode sont capables d'améliorer considérablement la précision d'apprentissage.

Le deep learning repose sur une combinaison d'algorithmes de machine learning qui utilisent plusieurs transformations non linéaires pour modéliser des abstractions de haut niveau dans les données.

Le deep learning est un sous-ensemble de l'apprentissage automatique et de l'IA. Il s'agit essentiellement d'un réseau neuronal à trois couches ou plus. Ces réseaux neuronaux visent à imiter l'activité du cerveau humain ; toutefois, ils sont loin d'être à la hauteur de la capacité du cerveau humain à apprendre à partir de grandes quantités de données. Si un réseau neuronal à une seule couche peut fournir des prédictions approximatives, des couches cachées supplémentaires peuvent aider à optimiser et à affiner la précision.

L'IA est déjà une réalité dans l'agriculture. Le marché de l'IA agricole était évalué à près de 518,7 millions de dollars en 2017 et devrait croître de plus de 22,5 % par an pour atteindre 2,6 milliards de dollars d'ici 2025. L'application et la

demande croissante de surveillance et d'analyse continues de la santé des cultures font progresser le marché des solutions d'intelligence artificielle basées sur la technologie de vision par ordinateur.

Les réseaux de neurones (profonds ou non) pour des cas agricoles sont explorés/utilisés de façon importante par les chercheurs. Le graphique ci-dessous, montre la part des réseaux de neurones par rapport aux autres algorithmes utilisés pour des cas agricoles **

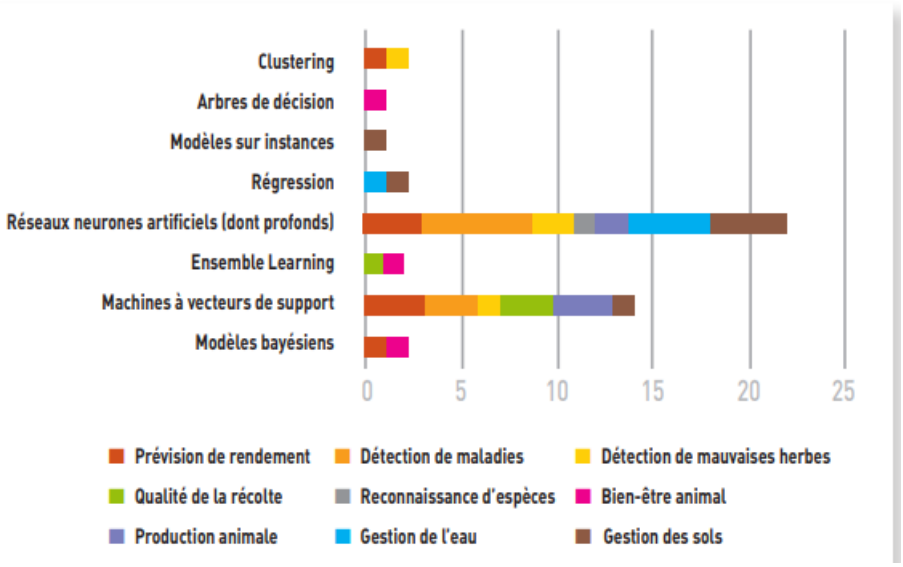


Figure 3:Nombre de modèles de machine learning pour des usages agricoles.

Convolutional Neural Network : Réseau neuronal convolutif (CNN)

Un CNN est un algorithme d'apprentissage profond composé de plusieurs couches convolutives, de couches de mise en commun et de couches entièrement connectées. Il s'agit d'un réseau neuronal multicouche basé sur le cortex visuel animal. Les CNN sont principalement utilisés pour le traitement d'images et la reconnaissance de caractères manuscrits. Les CNN ont été utilisés, entre autres, pour la classification d'images, la détection d'objets, la fragmentation d'images, la reconnaissance vocale, le traitement de textes et de vidéos et l'analyse d'images médicales. L'architecture d'un CNN se compose généralement de couches convolutives, de couches de mise en commun et de couches entièrement connectées.

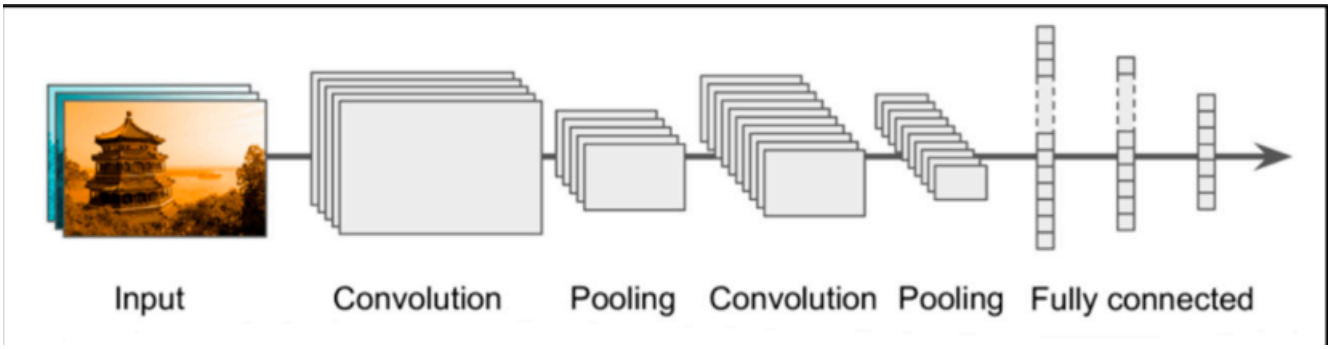


Figure 4: Architecture de CNN[1]

Recurrent Neural Network : Réseau neuronal récurrent (RNN)

Un RNN est un modèle de séquence de neurones qui donne des résultats exceptionnels dans des tâches cruciales telles que la modélisation du langage, la reconnaissance vocale et la traduction automatique. Contrairement aux réseaux de neurones traditionnels, les RNN tirent parti des informations séquentielles du réseau ; cet attribut est essentiel dans de nombreuses applications où la structure inhérente à la séquence de données contient des informations précieuses. Par exemple, pour comprendre un mot dans une phrase, il faut d'abord comprendre le contexte. Par conséquent, un RNN peut être considéré comme une unité de mémoire à court terme composée de la couche d'entrée x, de la couche cachée (d'état) s et de la couche de sortie y.

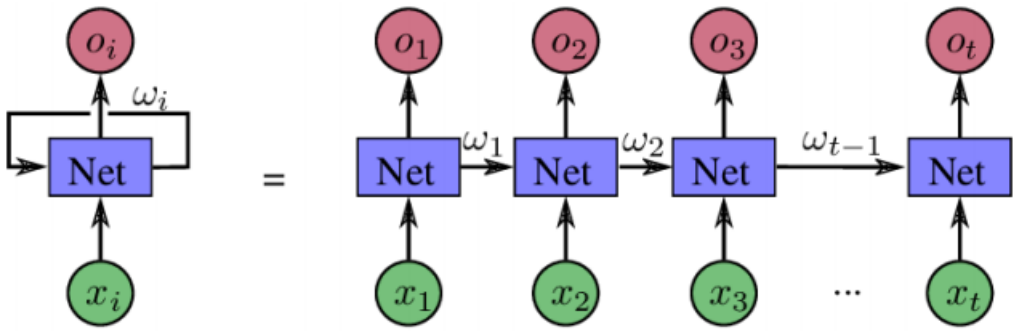


Figure 5: La structure générique du RNN[1]

Partie pratique

Méthodologie

Lors de la mise en œuvre du projet, les étapes suivantes ont été mises en œuvre afin d'atteindre les résultats

Nettoyage et prétraitement des données

L'une des premières étapes consiste à s'assurer que l'ensemble de données que nous utilisons est exact. L'ensemble de données ne doit pas comporter de valeurs manquantes et si c'est le cas, elles doivent être remplacées par les valeurs appropriées. Les données doivent également être vérifiées pour voir s'il existe une distribution normale de leurs caractéristiques. Les valeurs aberrantes doivent être supprimées. La valeur de l'asymétrie des caractéristiques doit être vérifiée et si les caractéristiques ont une asymétrie, alors ces caractéristiques doivent être normalisées en utilisant des transformations. L'ensemble de données que nous avons utilisé contenait des caractéristiques ayant une asymétrie. Pour les normaliser, nous avons utilisé une transformation quantile sur les caractéristiques de notre dataset.

Analyse et visualisation des données

Après avoir effectué le nettoyage et le prétraitement des données, nous procédons à l'analyse et à la visualisation de notre ensemble de données. Nous essayons d'analyser nos données plus clairement pour trouver des tendances ou des modèles dans l'ensemble de données. Nous avons créé plusieurs visualisations de notre ensemble de données afin de comprendre correctement les données. Nous avons créé des diagrammes à barres, des diagrammes de dispersion, des diagrammes en boîte, etc. afin de visualiser les données et de déterminer s'il existe des tendances ou des modèles qui nous seront utiles lors de la mise en œuvre de notre modèle.

Sélection des caractéristiques

Il est important que nous ne retenions que les caractéristiques qui seront nécessaires pour déterminer le type de culture à réaliser. Pour cela, nous avons créé une matrice de corrélation qui montre la relation linéaire d'une caractéristique avec toutes les autres caractéristiques. Si les caractéristiques sont fortement corrélées, elles doivent être abandonnées, mais comme nous pouvons le voir dans la matrice ci-dessous, les caractéristiques ne sont pas fortement corrélées entre elles, il est donc logique de n'en abandonner aucune et nous les utiliserons toutes pour prédire le type de culture.

Voici la matrice de corrélation pour nos caractéristiques dans notre dataset.

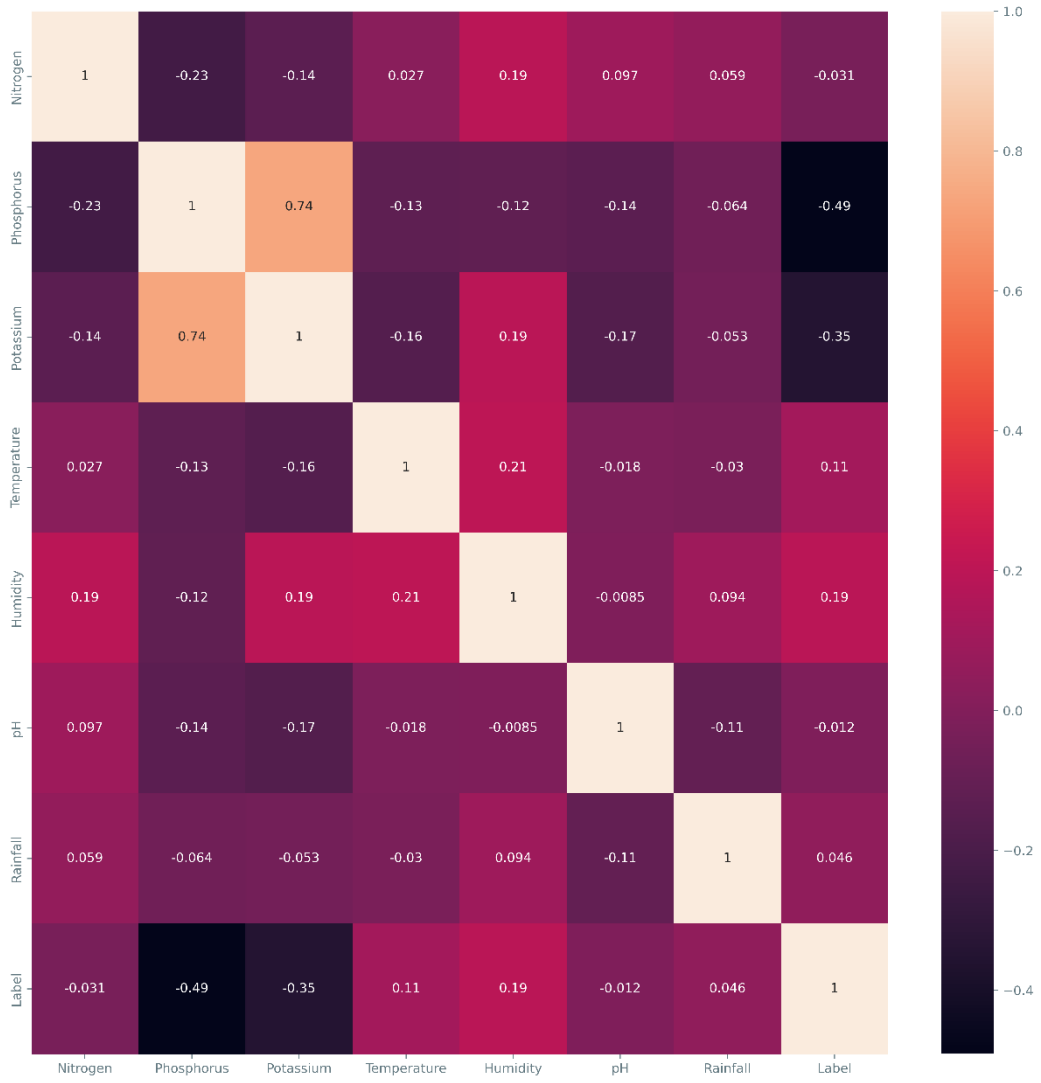


Figure 6:La matrice de corrélation pour nos caractéristiques dans notre dataset

Construction de modèles

L'étape suivante consiste à construire le modèle de machine learning. Lors de la construction du modèle d'apprentissage automatique, nous devons d'abord diviser notre ensemble de données en deux parties : training data et test data. Nous avons divisé les données dans un rapport de 80-20%. En prenant training set, nous appliquons nos algorithmes de machine learning sur les caractéristiques du dataset.

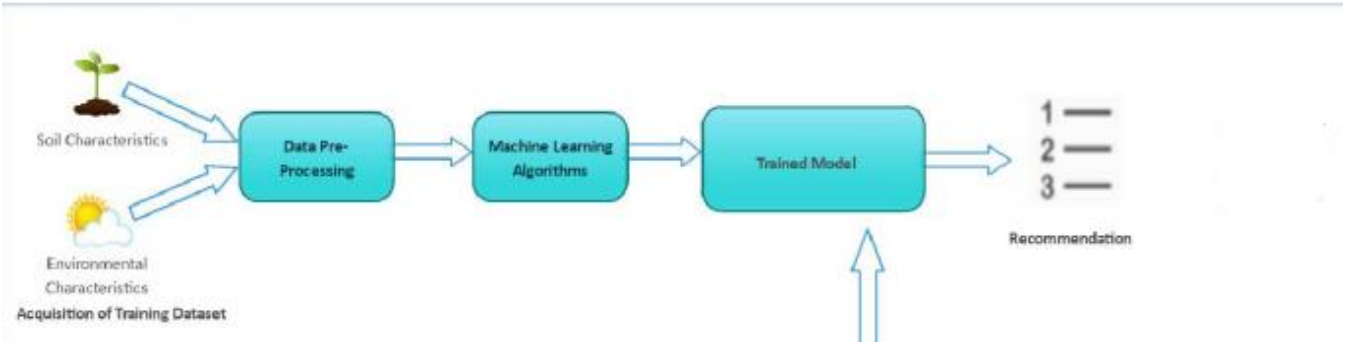


Figure 7:Processus de Machine Learning

Notre jeu de données :

La dataset pour ce sujet provient de Kaggle. Il s'agit d'un jeu de données de recommandation de cultures qui nous donne des informations sur les différents types de cultures et les caractéristiques qui décident de la culture appropriée.

Caractéristiques de ce jeu de données :

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Figure 8: Notre dataset

N : ratio d'Azote dans le sol

P : ratio de phosphore dans le sol

K : ratio Potassium dans le sol

Température : température en degrés Celsius

Humidité : humidité relative en %.

pH : valeur du pH du sol

rainfall : pluviométrie en mm

Algorithmes de machine learning utilisés

K-Nearest Neighbor (KNN): k plus proches voisins

L'algorithme K-Nest Neighbour est l'un des algorithmes d'apprentissage automatique les plus simples basés sur la technique de l'apprentissage supervisé. Il suppose la similitude entre le nouveau cas/données et les cas disponibles et place le nouveau cas dans la catégorie qui est la plus similaire aux catégories disponibles. Cela signifie que lorsqu'une nouvelle donnée apparaît, elle peut être facilement classée dans une catégorie bien définie en utilisant l'algorithme K-NN. Il peut être utilisé pour la régression comme pour la classification, mais il est surtout utilisé pour les problèmes de classification.

Régression logistique

C'est l'un des algorithmes les plus simples de l'apprentissage automatique. Il est utilisé pour résoudre les problèmes de classification. Il utilise une fonction sigmoïde pour calculer mathématiquement la probabilité d'une observation. D'une observation et, en conséquence, l'observation est alors placée dans sa classe respective. Lors du calcul, si la probabilité d'une observation est 0 ou 1, une valeur seuil est décidée et les classes les classes ayant des probabilités supérieures à la valeur seuil reçoivent la valeur seuil, la valeur 1 est attribuée aux classes ayant des seuils, la valeur 0.

Random Forest Classifier

Random Forest est un algorithme d'apprentissage automatique supervisé utilisé dans les problèmes de classification et de régression. Il contient plusieurs arbres de décision et une moyenne de ceux-ci est prise pour donner le résultat. Il est basé sur le concept de mise en sac dans lequel de multiples arbres de décision sont créés et une moyenne de ceux-ci est prise afin de donner la sortie. Comme les arbres de décision sont enclins à l'ajustement excessif, la forêt aléatoire est utile pour réduire l'effet de l'ajustement excessif et donc donner un résultat plus précis.

Light Gradient Boosting Machine

LightGBM est un modèle de boosting de gradient basé sur des arbres de décision pour augmenter l'efficacité du modèle et réduire l'utilisation de la mémoire : Il utilise deux nouvelles techniques : l'échantillonnage unilatéral basé sur le gradient et l'Exclusive Feature Bundling (EFB) qui répond aux limitations de l'algorithme basé sur l'histogramme qui est principalement utilisé dans tous les cadres GBDT (Gradient Boosting Decision Tree).

Réseau neuronal :

Les réseaux neuronaux sont un ensemble d'algorithmes, librement modelés sur le cerveau humain, qui sont conçus pour reconnaître des modèles. Ils interprètent les données sensorielles par une sorte de perception artificielle, en étiquetant ou en regroupant les données brutes. Les modèles qu'ils reconnaissent sont numériques, contenus dans des vecteurs, dans lesquels toutes les données du monde réel, qu'il s'agisse d'images, de sons, de textes ou de séries chronologiques, doivent être traduites. Les réseaux neuronaux sont eux-mêmes des approximations de fonctions générales, c'est pourquoi ils peuvent être appliqués à presque tous les problèmes d'apprentissage automatique concernant

l'apprentissage d'une correspondance complexe entre l'espace d'entrée et l'espace de sortie.

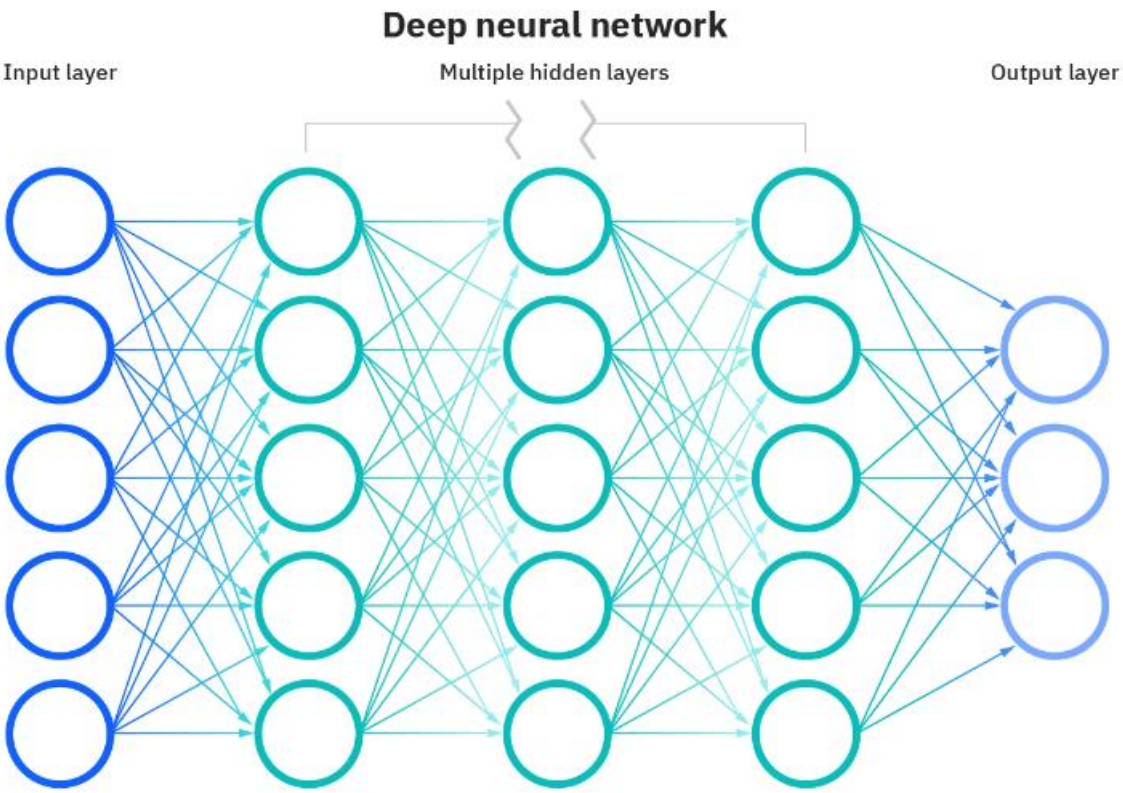


Figure 9: Deep neural network

La solution réalisée

Outils et environnement d'exécution

Python :

Python est un choix extrêmement populaire dans le domaine de l'apprentissage automatique et du développement de l'IA. Sa syntaxe courte et simple le rend extrêmement facile à apprendre.

Google Colab :

Google Colab ou Colaboratory est un service cloud, offert par Google (gratuit), basé sur Jupyter Notebook et destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet d'entraîner des modèles de Machine Learning directement dans le cloud. Colab permet :

- De développer des applications en Deep Learning en utilisant des bibliothèques
  Python populaires telles que Keras, TensorFlow, PyTorch et OpenCV ;
- D'utiliser un environnement de développement (Jupyter Notebook) qui ne nécessite aucune configuration…

Mais la fonctionnalité qui distingue Colab des autres services est l'accès à un processeur graphique GPU (et Tensor Processing Unit TPU), totalement gratuitement.

Librairies :

- Pandas : une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse des données.
- NumPy : une extension du langage de programmation Python, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux.
- Matplotlib : une bibliothèque du langage de programmation Python destinée à tracer et visualiser des données sous formes de graphiques.
- Seaborn : est une bibliothèque de visualisation de données en Python basée sur matplotlib. Elle fournit une interface de haut niveau pour dessiner des graphiques statistiques attrayants et informatifs.
- Scikit-learn : est une bibliothèque d'apprentissage automatique qui permet l'apprentissage supervisé et non supervisé. Des données, la sélection et l'évaluation des modèles.
- Keras : est une bibliothèque open source écrite en python permet d'interagir avec les algorithmes de réseaux de neurones profonds et de machine learning,

conçue pour permettre une expérimentation rapide avec les réseaux de neurones profonds.

❖ TensorFlow est une bibliothèque open-source développée par Google principalement pour les applications d'apprentissage profond. Elle prend également en charge l'apprentissage automatique traditionnel.

La réalisation :

Acquisition des données :

```
[ ]  from google.colab import drive
     drive.mount('/content/drive')

     Mounted at /content/drive

[ ]  df=pd.read_csv('/content/drive/MyDrive/PFA/Crop_recommendation.csv', sep=",")
```

Figure 10:Acquisition des données

Nettoyage et prétraitement des données :

```
[ ]  #checking the missing values
     df.isnull().sum()

     N              0
     P              0
     K              0
     temperature    0
     humidity       0
     ph             0
     rainfall       0
     label          0
     dtype: int64
```

```
[ ]  df['label'].unique()

     array(['rice', 'maize', 'chickpea', 'kidneybeans', 'pigeonpeas',
            'mothbeans', 'mungbean', 'blackgram', 'lentil', 'pomegranate',
            'banana', 'mango', 'grapes', 'watermelon', 'muskmelon', 'apple',
            'orange', 'papaya', 'coconut', 'cotton', 'jute', 'coffee'],
           dtype=object)
```

Figure 11: Prétraitement des données

Analyse et visualisation des données :

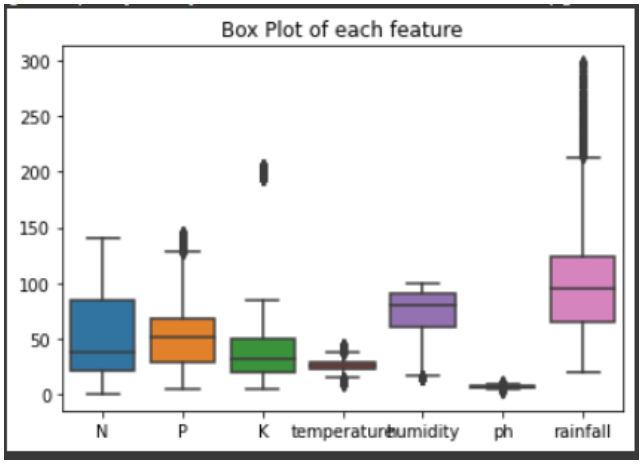


Figure 12:Boîte à moustaches

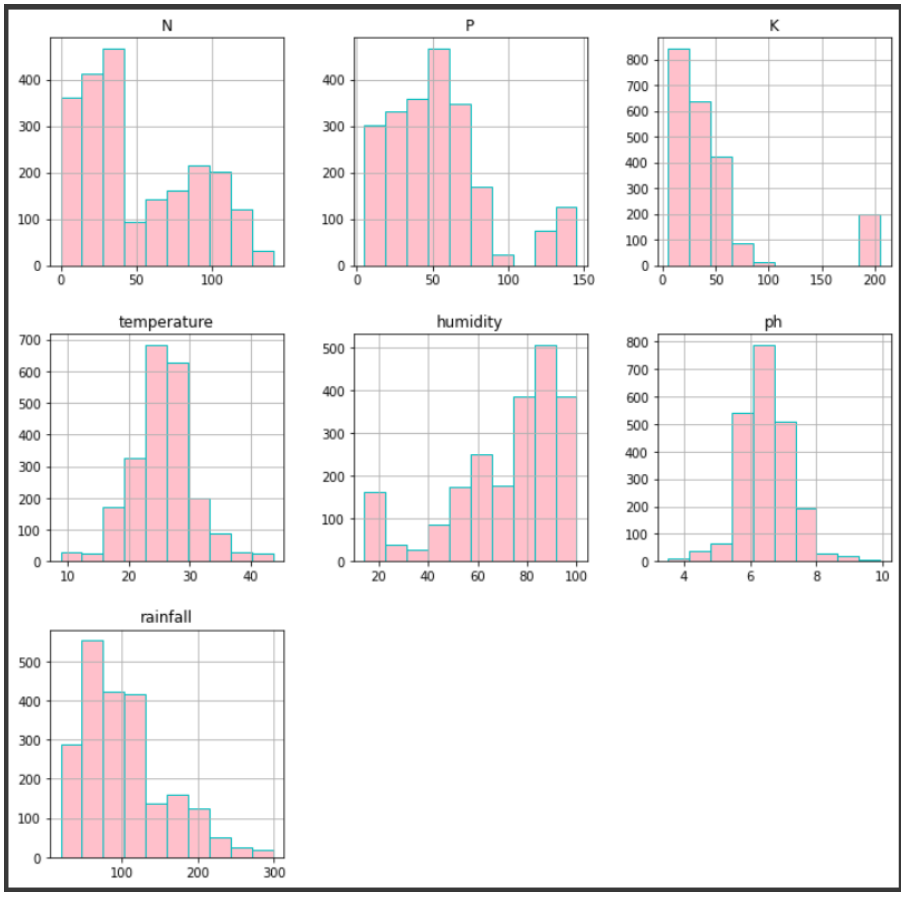


Figure 13:Visualisation des variables

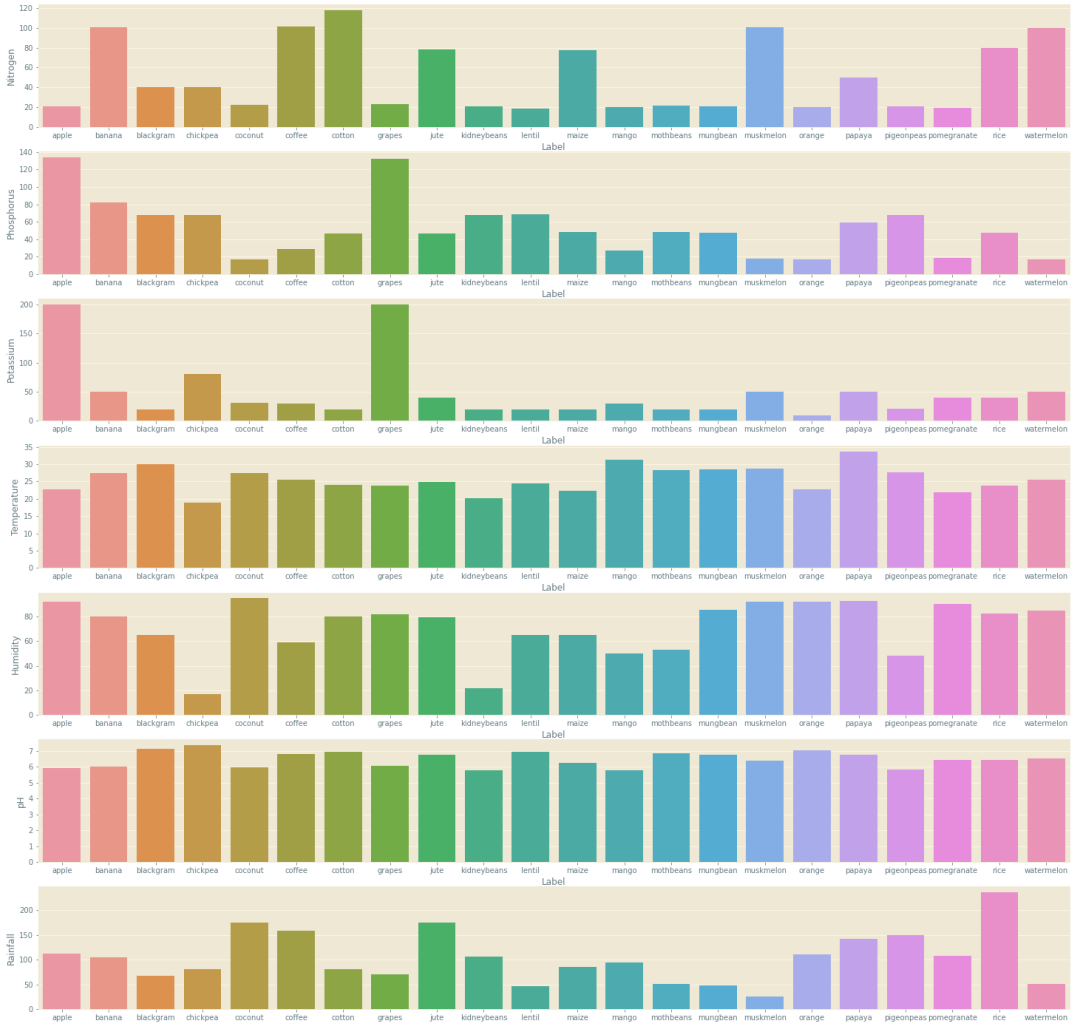


Figure 14:Comparaison des attributs moyens de différentes classes

Observations :

- Le coton nécessite le plus d'azote.
- La pomme a besoin le plus de phosphore.
- Le raisin a besoin le plus de potassium.
- La papaye exige un climat chaud.
- Coconut exige un climat humide.
- Le pois chiche nécessite un pH élevé dans le sol.
- Le riz nécessite d'importantes précipitations.

Corrélation entre les variables :

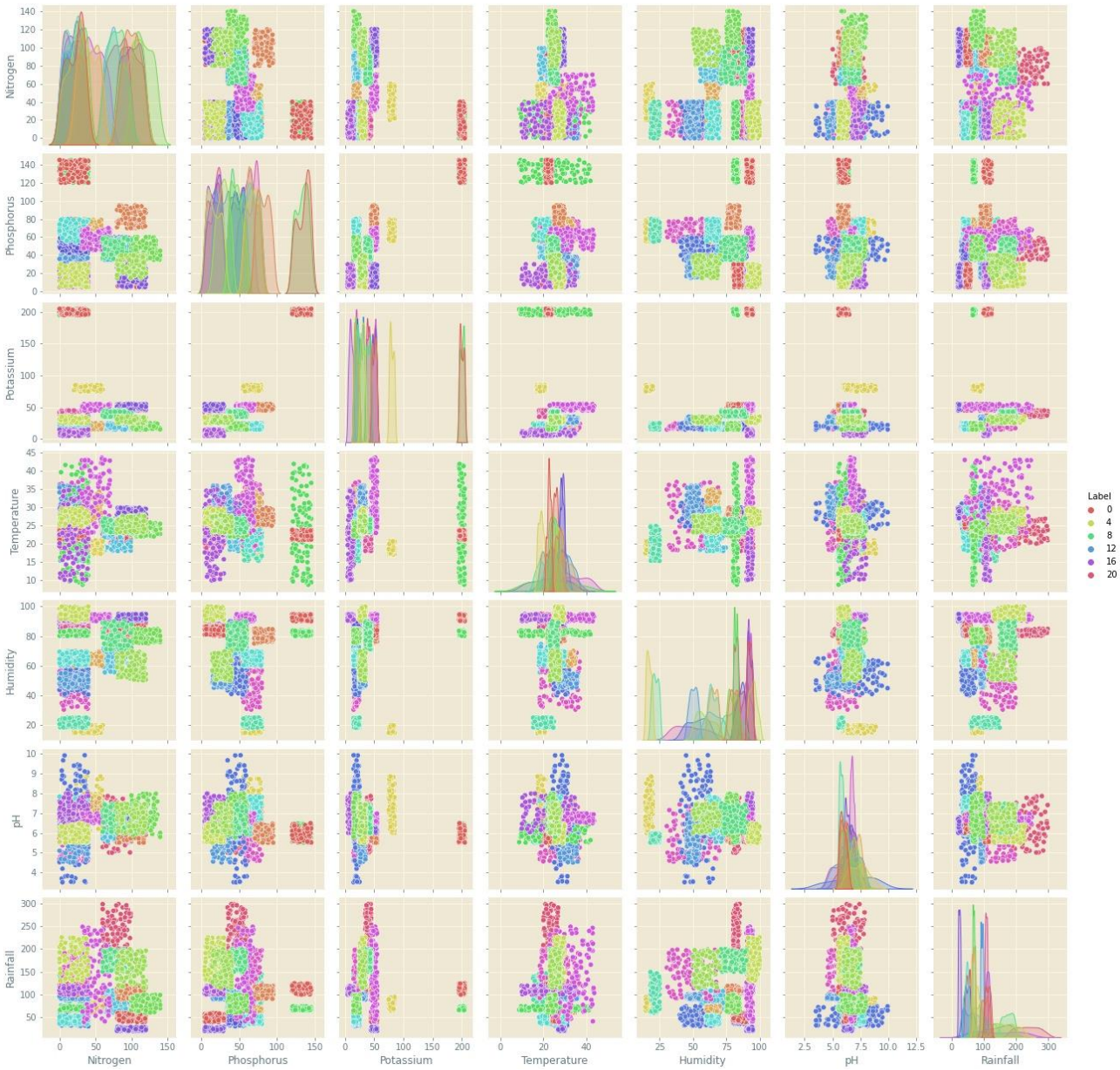


Figure 15: Visualisation des classes par pairplot

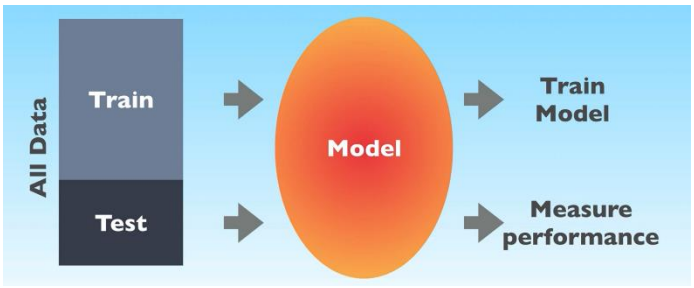
Encodage

```
[ ]  names = df['Label'].unique()

[ ]  names

     array(['rice', 'maize', 'chickpea', 'kidneybeans', 'pigeonpeas',
            'mothbeans', 'mungbean', 'blackgram', 'lentil', 'pomegranate',
            'banana', 'mango', 'grapes', 'watermelon', 'muskmelon', 'apple',
            'orange', 'papaya', 'coconut', 'cotton', 'jute', 'coffee'],
           dtype=object)

     from sklearn.preprocessing import LabelEncoder
     encoder=LabelEncoder()
     df['Label']=encoder.fit_transform(df['Label'])
     df
```

Figure 16:Encodage

Ici, nous avons encodé les valeurs cibles dans leur valeur numérique respective. C'est nécessaire parce que notre modèle de machine learning ne sera pas capable de comprendre les chaînes de caractères !

Répartition des données en trainig and test set




```
[ ]  #Split into training and test set
     from sklearn.model_selection import train_test_split
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2,shuffle = True,stratify=y)
```

- test_size : Indique quelle proportion du total des échantillons doit être donnée à test set - ici 20%.
- shuffle : La dataset contient toutes les classes et ses échantillons un par un. Il est donc nécessaire de la mélanger pour éviter tout biais.
- stratify : Assure une distribution égale des classes entre train et test set.

Figure 17:Train/test split

Construction de modèles

Ce travail, présente 6 Algorithmes, dont 4 ML (KNN, RF, Logistic Regression, LGBM) et un deep learning (Fully Connected Neural Network).

K-Nearest Neighbor (KNN):

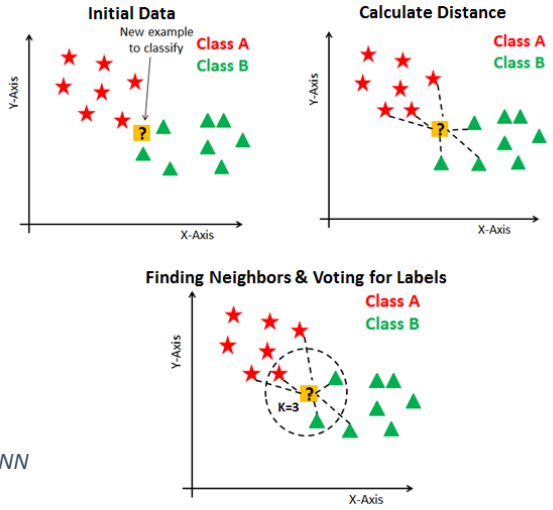


Figure 18:Processus de KNN

Confusion Matrix

La matrice de confusion s'agit d'une mesure de performance pour un problème de classification par machine learning où la sortie peut être deux classes ou plus. C'est un tableau avec 4 combinaisons différentes de valeurs prédites et réelles.

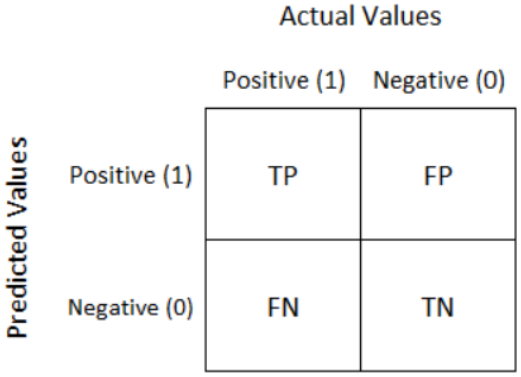


Figure 19: Matrice de confusion

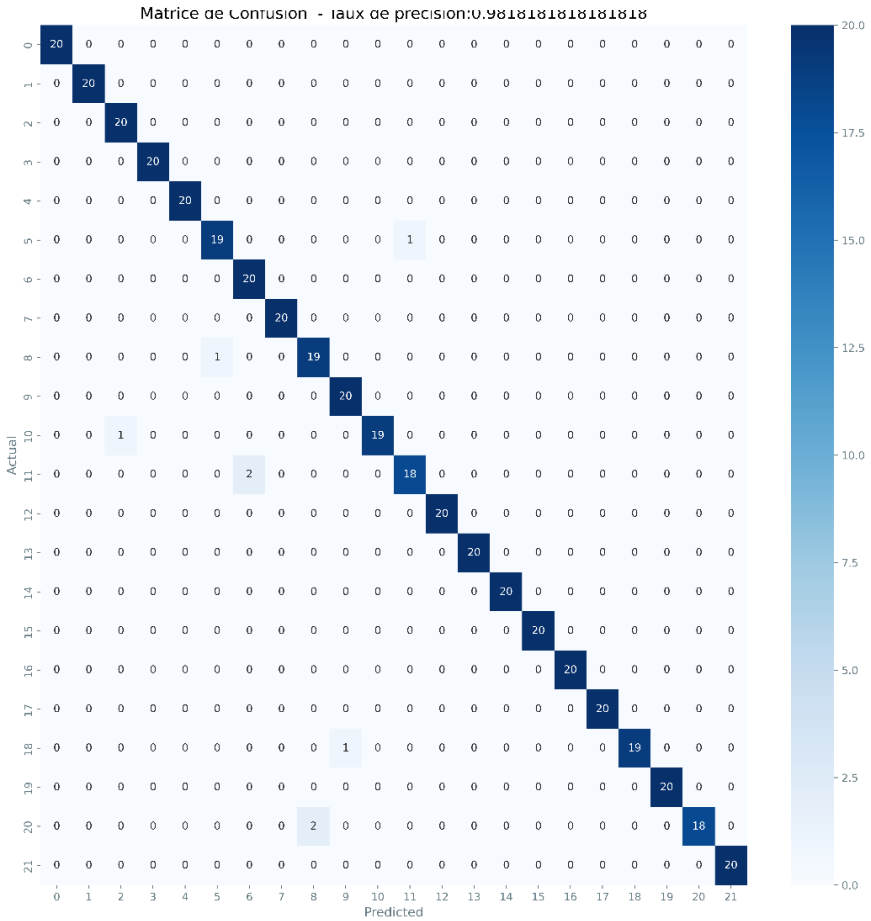


Figure 20: Matrice de confusion par le KNN

Régression logistique :

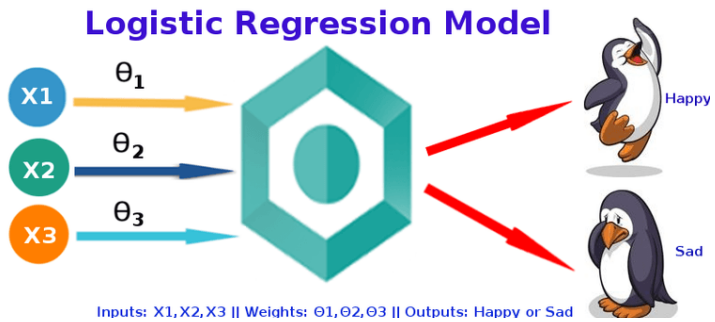


Figure 21: Démarche de la régression logistique

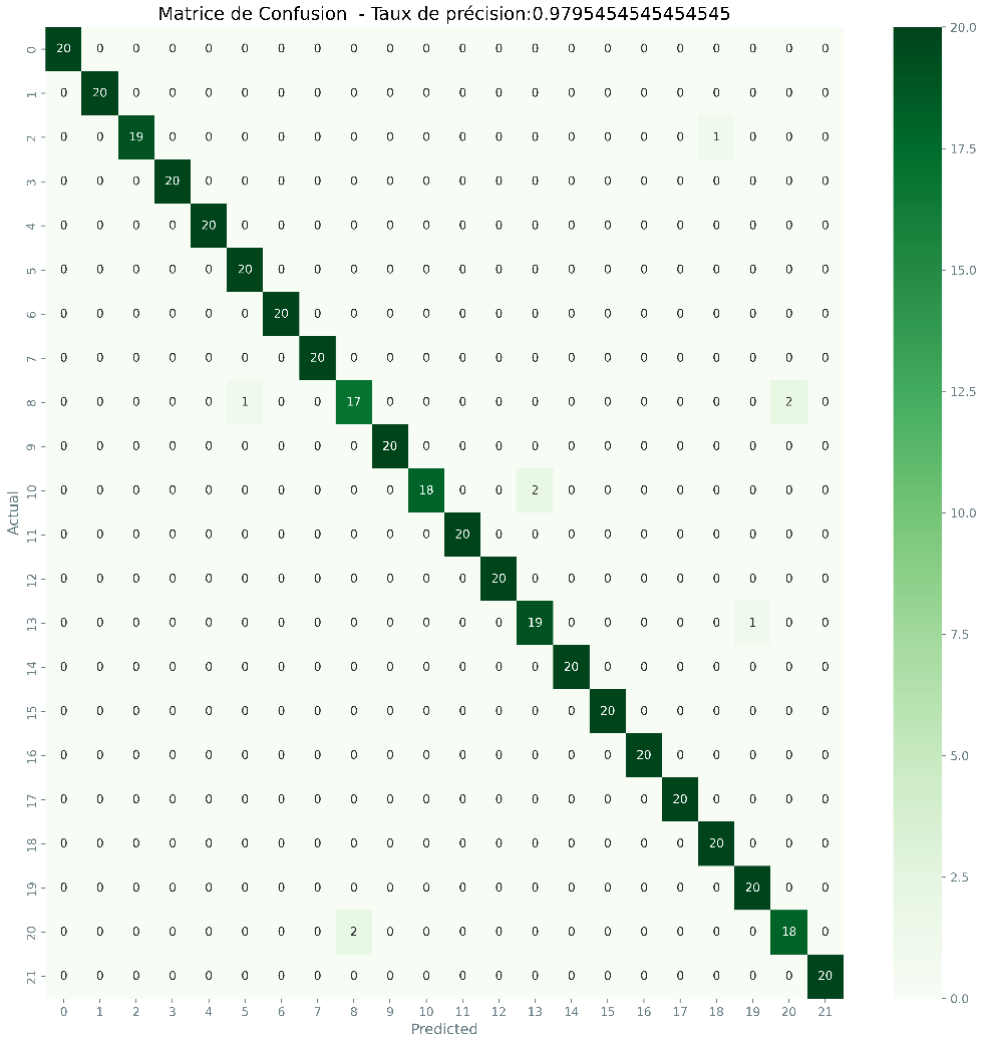


Figure 22:Matrice de confusion par la régression logistique

Random Forest Classifier

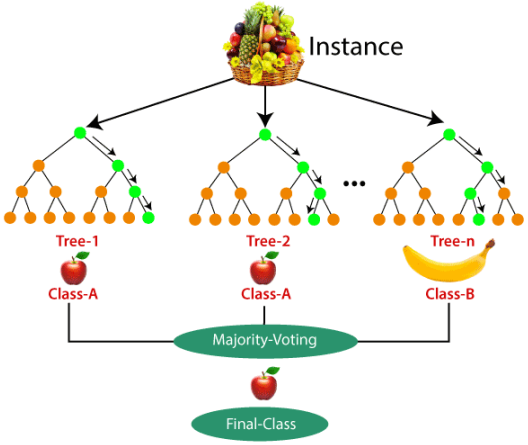


Figure 23Random Forest Classifier

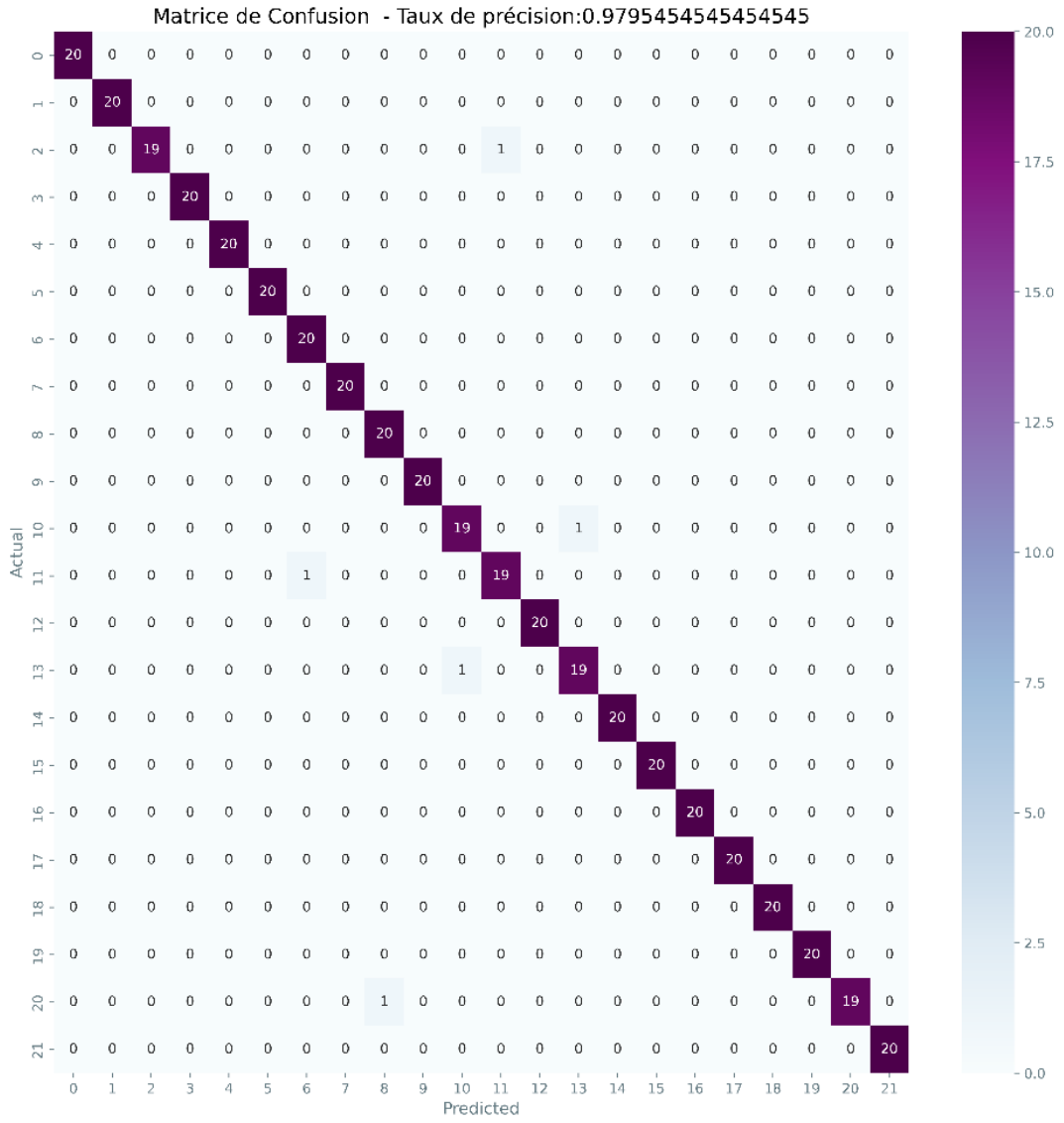


Figure 24:Matrice de confusion par le Random Forest Classifier

Light Gradient Boosting Machine :

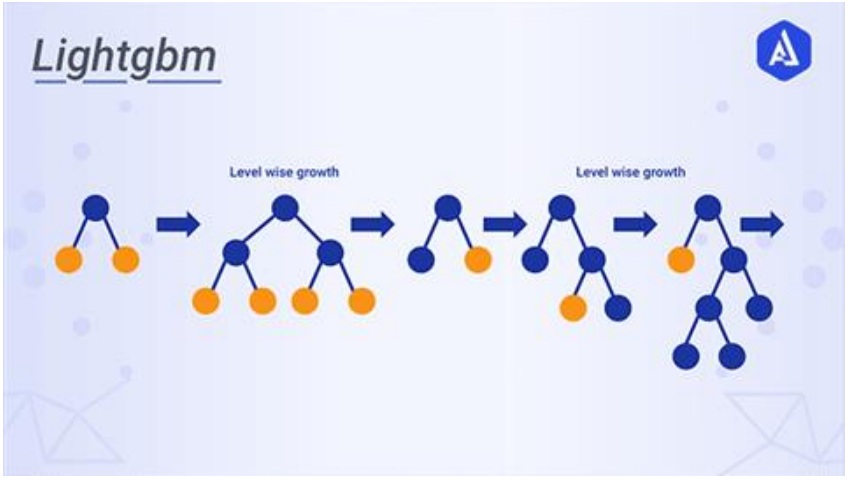


Figure 25:Light Gradient Boosting Machine

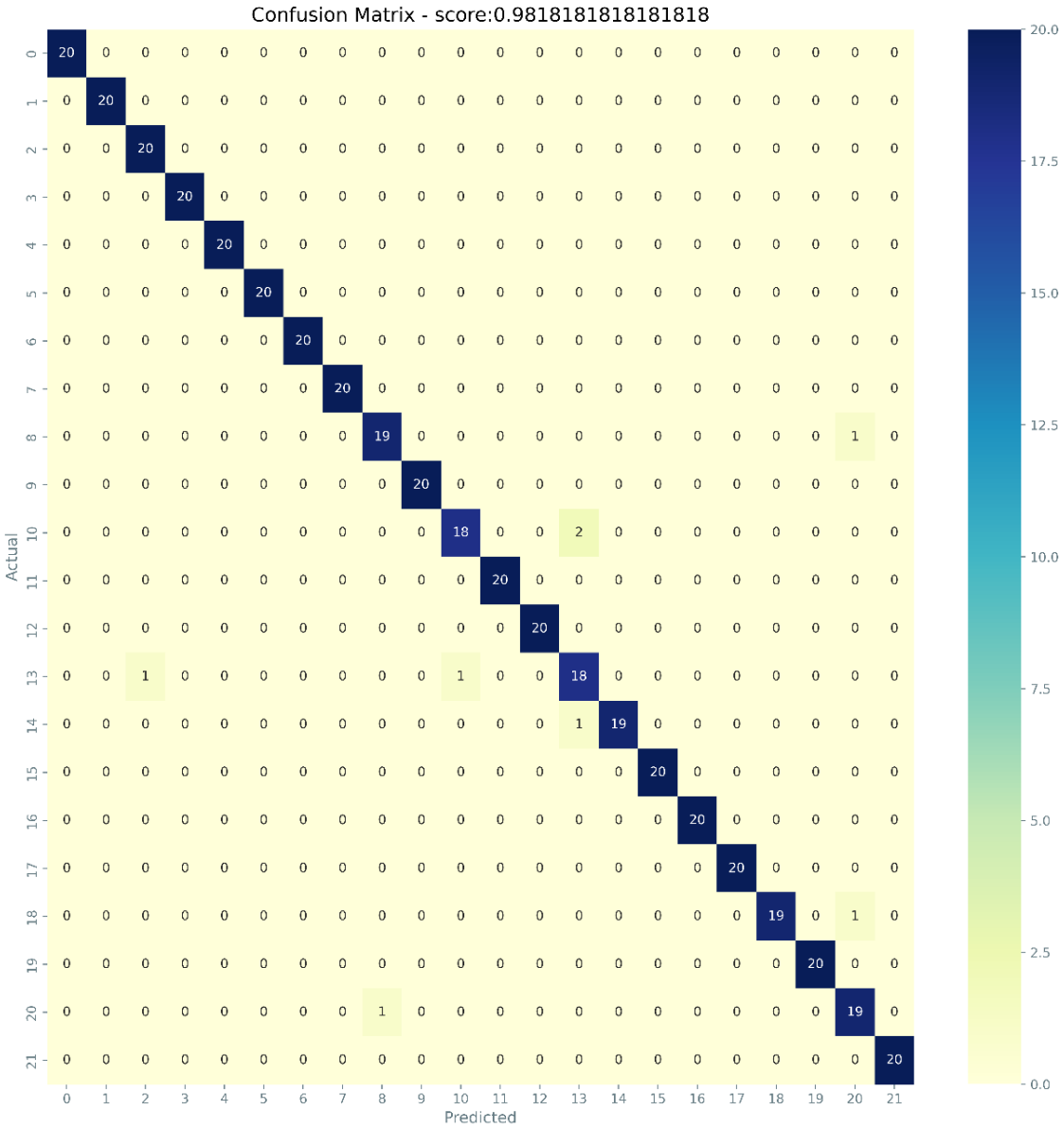


Figure 26:Matrice de confusion par le LGBM

Fully Connected Neural Network :

```
import tensorflow.keras.backend as K
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.optimizers import Adam
```

```
model = Sequential()
K.clear_session()
model.add(Dense(8, activation='relu', input_shape=(7,)))
model.add(Dense(16, activation='relu',))
model.add(Dense(32, activation='relu'))
model.add(Dense(22, activation='softmax'))
model.compile(optimizer='adam',loss = tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True), metrics=['accuracy'])
```

```
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 8)                 64

 dense_1 (Dense)             (None, 16)                144

 dense_2 (Dense)             (None, 32)                544

 dense_3 (Dense)             (None, 22)                726

=================================================================
Total params: 1,478
Trainable params: 1,478
Non-trainable params: 0
_________________________________________________________________
```

```
model.fit(X_train, y_train,epochs=100, batch_size=64, verbose=1)
```

Si vous avez remarqué, nous avons utilisé la fonction *ReLu* comme fonction d'activation,

- ReLU (*Rectified Linear Units*) désigne la fonction réelle non-linéaire définie par : $ReLu(x)$ = max (0, x).
  Elle remplace donc toutes les valeurs négatives reçues en entrées par des zéros.

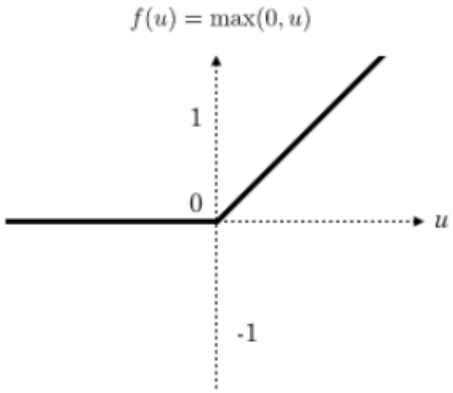


Figure 27: Courbe de la fonction d'activation ReLu

- Aussi,pour l'output la fonction Softmax, cette variante de softmax calcule la probabilité de chaque classe possible. Nous l'utiliserons le plus lorsqu'il s'agira de traiter des réseaux de neurones multiclasses en Python.

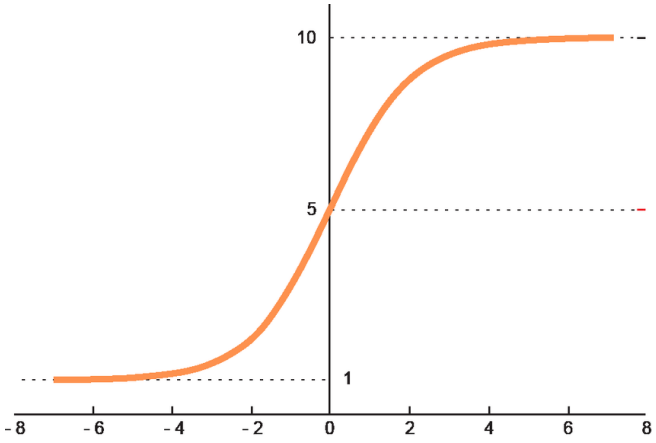


Figure 28:Courbe de la fonction d'activation Softmax

- L'optimizer Adam est une méthode de descente de gradient stochastique qui repose sur l'estimation adaptative des moments de premier et de second ordre.
- La categorical cross-entropy loss est également connue sous le nom de log-vraisemblance négative. Il s'agit d'une fonction de perte populaire pour les problèmes de catégorisation et elle mesure la similarité entre deux distributions de probabilité, généralement les étiquettes réelles et les étiquettes prédites.

Nous constatons qu'ils dépassent tous les 90% donc performants, nous avons choisis de prédire les meilleures cultures selon le modèle du LightGBM :

✅ | Prévision des Top 3

```
[ ]  Semences=df['label'].unique()
```

```
[ ]  Semences.sort()
```

```
[ ]  def predit(X):
         probability=model.predict_proba(X)
         probability = sorted( [(x,i) for (i,x) in enumerate(probability[0])], reverse=True)
         for i,j in probability[:3]:
             print(Semences[j])
     predit(X_test.sample(1))
```

```
orange
pomegranate
rice
```

Figure 29: Prévision des Top 3

Conclusion

Ce stage fut une expérience très enrichissante pour le nombre de connaissances que j'ai acquise, mais aussi pour l'extension de mon réseau relationnel.

Nous avons traite un sujet très pertinent et récent qu'est le Smart Farming ou l'Agriculture Intelligente, et donc qui nous n'était pas sans contraintes. Des obstacles que nous nous sommes réjouis de surpasser ensemble avec l'aide et la bienveillance de l'encadrant Monsieur Chouaib Moujahdi.

Le travail était d'autant plus intéressant, comme il était en binôme, j'ai eu la chance de mettre à profit mon esprit d'équipe qui s'est révélé encore une fois fructueux.

Il en convient de dire, que ce stage était surtout une affirmation ou confirmation de ma volonté de poursuivre dans ce domaine.

Références

Http://www.israbat.ac.ma

[Ünal 2020] Ünal, Zeynep. (2020). Smart Farming Becomes Even Smarter with Deep Learning—A Bibliographical Analysis. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3000175.

[Altalak et al. 2022] Altalak, Maha & uddin, Mohammad & Alajmi, Amal & Rizg, Alwaseemah. (2022). Smart Agriculture Applications Using Deep Learning Technologies: A Survey. Applied Sciences. 12. 5919. 10.3390/app12125919.

Jain, Virbahu. (2019). Robotics for Supply Chain and Manufacturing Industries and Future It Holds! International Journal of Engineering Research and. V8. 10.17577/IJERTV8IS030062.

Kaur, Harjeet & Prashar, Deepak & Madhuri, (2019). Deep Learning in Smart Farming - Concepts, Applications and Techniques. 3359.

Memon, Muhammad & Kumar, Pardeep & Mirani, Azeem & Qabulio, Mumtaz & Sodhar, Irum Hafeez. (2020). Deep Learning and IoT: The Enabling Technologies Towards Smart Farming.10.4018/978-1-7998-2803-7.ch003.
https://www.agrotic.org/wpcontent/uploads/2018/12/2018_ChaireAgroTIC_DeepLearning_VD2.pdf

https://www.college-de-france.fr/site/yann-lecun/course-2015-2016.htm.

Kiranyaz, Serkan & Gastli, Adel & Ben-Brahim, L. & Alemadi, Nasser & Gabbouj, Moncef. (2018). Real-Time Fault Detection and Identification for MMC Using 1-D Convolutional Neural Networks. IEEE Transactions on Industrial Electronics. PP. 1-1. 10.1109/TIE.2018.2833045.

Wijerathna Yapa, Akila. (2022). Gains From Smart Farming. 114.

Durai, Senthil & Shamili, Mary. (2022). Smart farming using Machine Learning and Deep Learning techniques. Decision Analytics Journal. 3. 100041. 10.1016/j.dajour.2022.100041.

Heera, S.S. & Rahul, D. & Athreya, S.S. & Narasimman, S.S. & Harini, K.M. (2020). Smart farming using deep learning technique. International Journal of Control and Automation. 13.771-775.