



Réaliser par :

Lkamel Mohamed

Oublal Mohamed

Farah Oussama

ETTAFSSAOUI YOUSSEF

Encadrée par :

ACHARGVI ZINEB

Remerciement

Au nom d'Allah le tout miséricordieux, le très miséricordieux.

Ce travail, ainsi accompli, n'aurait pas pu arriver terme sans l'aide et le soutien d'Allah, louange au tout miséricordieux, le seigneur de l'univers. Nous remercions en premier lieu nos parents qui ne préservent aucun effort pour nous voir escalader à pas sur la montagne du savoir, et dépasser tous les obstacles vers l'amélioration. Tout mot dit, nous ne les remercierons jamais assez. Nous tenons à remercier madame ACHARQUI ZINEB

de cette opportunité qui nous permet d'étudier un cas réel. Aussi pour ces conseils accordés tout dès le début de notre formation Après, nous tenons à adresser notre sincère remerciement ainsi que nos sentiments les plus respectueux à qui a mis à notre disposition toutes les ressources nécessaires, et qui n'a jamais hésité un instant nous orienter avec ses précieuses directives et ses judicieux conseils.

Scikit learn & Python

Introduction	3
Machine Learning	4
Définition :	4
Types d'apprentissage machine :	5
Utilisations de l'apprentissage machine :	7
Techniques d'apprentissage automatique :	8
Avantages et inconvénients de l'apprentissage machine :	9
Langages de programmation pour l'apprentissage machine :	10
Python.....	11
Définition :	11
A quoi sert le langage Python ?	11
Scikit-learn	13
Définition :	13
Installation de scikit-learn :	13
Application :	14
Conclusion.....	23

Introduction

L'apprentissage machine est un sous-domaine de l'intelligence artificielle (IA). L'objectif de l'apprentissage machine consiste généralement à comprendre la structure des données et à que les données en modèles qui peuvent être compris et utilisés par les gens.

Bien que l'apprentissage machine soit un domaine de l'informatique, il diffère à partir d'approches informatiques traditionnelles. Dans l'informatique traditionnelle, Les algorithmes sont des ensembles d'instructions explicitement programmées utilisées par des ordinateurs pour calculer ou résoudre des problèmes. Algorithmes d'apprentissage machine permettre aux ordinateurs de se former à la saisie de données et à l'utilisation des statistiques afin de produire des valeurs qui se situent dans une fourchette spécifique.

De ce fait, l'apprentissage machine facilite la construction d'ordinateurs des modèles à partir de données d'échantillons afin d'automatiser la prise de décision des processus basés sur l'entrée de données.

Tout utilisateur de technologie aujourd'hui a bénéficié de l'apprentissage machine.

La technologie de reconnaissance faciale permet aux plateformes de médias sociaux d'aider les utilisateurs taguer et partager des photos d'amis. Reconnaissance optique de caractères (OCR)

La technologie convertit les images de texte en caractères mobiles. Recommandation moteurs, alimentés par l'apprentissage machine, suggèrent quels films ou les émissions de télévision à regarder ensuite en fonction des préférences de l'utilisateur. L'auto-conduite Les voitures qui reposent sur l'apprentissage de la navigation par machine pourraient bientôt être disponibles pour consommateurs.

L'apprentissage machine est un domaine en constante évolution. C'est pourquoi, il y a quelques considérations à garder à l'esprit lorsque vous travaillez avec une machine les méthodologies d'apprentissage, ou analyser l'impact de l'apprentissage machine processus.

Machine Learning

Comportement. Seules les choses non vivantes sont les choses naturelles qui ne sont pas impliquées dans l'apprentissage. Il semble donc que vivre et apprendre vont de pair. Dans les choses non vivantes fabriquées par la nature, il n'y a presque rien à apprendre. Pouvons-nous introduire l'apprentissage dans les choses non vivantes fabriquées par l'homme et appelées machines ? Permettre à une machine capable d'apprendre comme les humains est un rêve, dont la réalisation peut nous amener à avoir des machines déterministes avec liberté (ou l'illusion de la liberté dans un sens). Pendant ce temps, nous pourrions nous vanter avec bonheur que nos humanoïdes ressemblent à l'image et à la ressemblance des humains sous forme de machines.

Définition :

L'apprentissage machine est un sous-domaine de l'intelligence artificielle, le terme faisant référence à la capacité des systèmes informatiques à trouver de manière indépendante des solutions aux problèmes en reconnaissant des modèles dans les bases de données. En d'autres termes, il s'agit de la capacité des systèmes informatiques à trouver de manière autonome des solutions à des problèmes en reconnaissant des modèles dans des bases de données : L'apprentissage machine permet aux systèmes informatiques de reconnaître des modèles sur la base d'algorithmes et d'ensembles de données existants et de développer des concepts de solutions adéquats. Par conséquent, dans l'apprentissage automatique, des connaissances artificielles sont générées sur la base de l'expérience.

Types d'apprentissage machine :

Fondamentalement, les algorithmes jouent un rôle important dans l'apprentissage automatique : D'une part, ils sont responsables de la reconnaissance de modèles et, d'autre part, ils peuvent générer des solutions. Les algorithmes peuvent être divisés en différentes catégories :

L'apprentissage supervisé : Au cours de l'apprentissage supervisé, des exemples de modèles sont définis à l'avance. Afin de garantir une attribution adéquate des informations aux groupes de modèles respectifs des algorithmes, ceux-ci doivent ensuite être spécifiés. En d'autres termes, le système apprend sur la base de paires d'entrées et de sorties données. Au cours de l'apprentissage contrôlé, un programmeur, qui agit comme une sorte d'enseignant, fournit les valeurs appropriées pour une entrée particulière. L'objectif est de former le système dans le contexte de calculs successifs avec différentes entrées et sorties et d'établir des connexions.

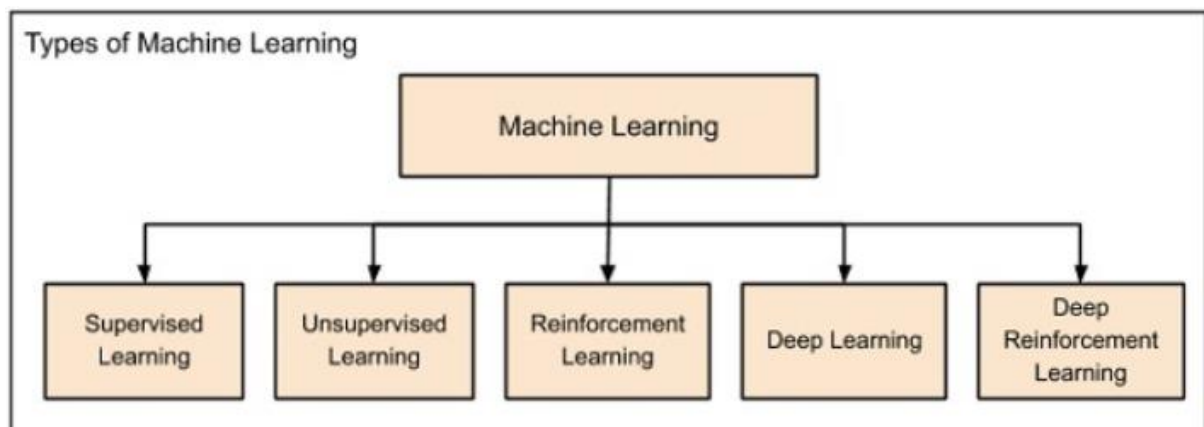
Apprentissage non supervisé : dans l'apprentissage non supervisé, l'intelligence artificielle apprend sans valeurs cibles prédéfinies et sans récompenses. Elle est principalement utilisée pour l'apprentissage de la segmentation (clustering). La machine tente de structurer et de trier les données saisies en fonction de certaines caractéristiques. Par exemple, une machine pourrait (très simplement) apprendre que des pièces de différentes couleurs peuvent être triées selon la caractéristique "couleur" afin de les structurer.

Apprentissage partiellement supervisé : l'apprentissage partiellement supervisé est une combinaison d'apprentissage supervisé et non supervisé.

Encourager l'apprentissage : Le renforcement de l'apprentissage - tout comme le conditionnement classique de Skinner - est basé sur des récompenses et des punitions. L'algorithme est enseigné par une interaction positive ou négative dont la réaction à une certaine situation doit avoir lieu.

L'apprentissage actif : Dans le cadre de l'apprentissage actif, un algorithme a la possibilité d'interroger les résultats pour des données d'entrée spécifiques sur la base de questions prédéfinies qui sont considérées comme significatives. Habituellement, l'algorithme lui-même sélectionne les questions les plus pertinentes.

En général, la base de données peut être soit hors ligne, soit en ligne, selon le système correspondant. En outre, elle peut être disponible une seule fois ou de manière répétée pour l'apprentissage automatique. Une autre caractéristique distinctive est soit le développement échelonné des paires d'entrée et de sortie, soit leur présence simultanée. Sur la base de cet aspect, une distinction est faite entre l'apprentissage dit séquentiel et l'apprentissage dit par lots.



Utilisations de l'apprentissage machine :

Les applications de l'apprentissage machine sont illimitées et de nombreux algorithmes d'apprentissage machine sont disponibles pour apprendre. Ils sont disponibles sous toutes les formes, de la plus simple à la plus complexe. Les 10 principales utilisations de l'apprentissage machine sont les suivantes :

Reconnaissance d'images : La reconnaissance d'images est l'une des utilisations les plus courantes des applications d'apprentissage machine. Elle peut également être appelée image numérique et pour ces images, la mesure décrit la sortie de chaque pixel d'une image. La reconnaissance des visages est également l'une des grandes caractéristiques qui ont été développées par l'apprentissage automatique uniquement. Elle permet de reconnaître le visage et d'envoyer les notifications qui s'y rapportent aux personnes.

Reconnaissance vocale : L'apprentissage machine (ML) aide également à développer l'application de reconnaissance vocale. Il est également appelé "assistant personnel virtuel" (APV). Il vous aidera à trouver les informations demandées par la voix. Après votre question, cet assistant recherchera les données ou les informations que vous avez posées et recueillera les informations nécessaires pour vous fournir la meilleure réponse. Il existe de nombreux dispositifs dans le monde actuel de l'apprentissage automatique pour la reconnaissance vocale, à savoir l'écho Amazon et les googles à la maison pour les locuteurs intelligents. Il existe une application mobile appelée Google allo et les smartphones sont les Samsung S8 et Bixby.

Prédictions : Il aide à construire les applications qui permettent de prévoir le prix d'un taxi ou d'un voyage pour une durée donnée et les embouteillages où l'on peut se trouver. Lors de la réservation du taxi et l'application estime le prix approximatif du voyage qui est fait par les utilisations de l'apprentissage machine seulement. Lorsque nous utilisons le service GPS pour vérifier l'itinéraire de la source à la destination, l'application nous montre les différentes façons d'aller vérifier le trafic à ce moment pour le nombre de véhicules le plus faible et où l'encombrement du trafic est plus que ce qui est fait ou récupéré par les utilisations de l'application d'apprentissage automatique.

Techniques d'apprentissage automatique :

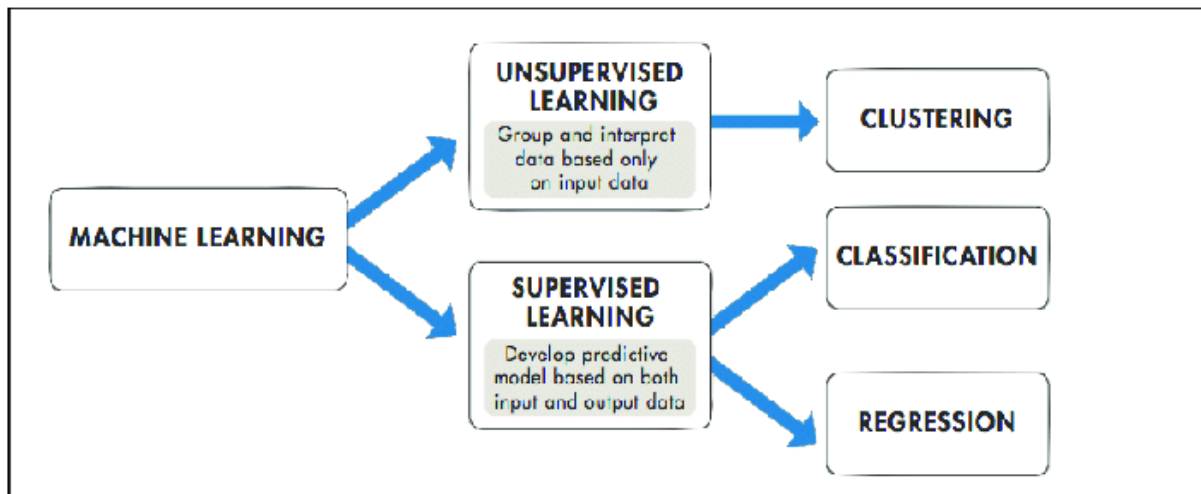
Il existe quelques méthodes qui ont une influence sur la promotion des systèmes pour apprendre et s'améliorer automatiquement en fonction de l'expérience. Mais elles relèvent de diverses catégories ou types comme l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage de renforcement, l'apprentissage par représentation, etc. Vous trouverez ci-dessous les techniques qui relèvent de l'apprentissage automatique :

Régression : Les algorithmes de régression sont principalement utilisés pour faire des prévisions sur les nombres, c'est-à-dire lorsque la sortie est une valeur réelle ou continue. Comme il relève de l'apprentissage supervisé, il fonctionne avec des données formées pour prédire de nouvelles données d'essai. Par exemple, l'âge peut être une valeur continue puisqu'il augmente avec le temps.

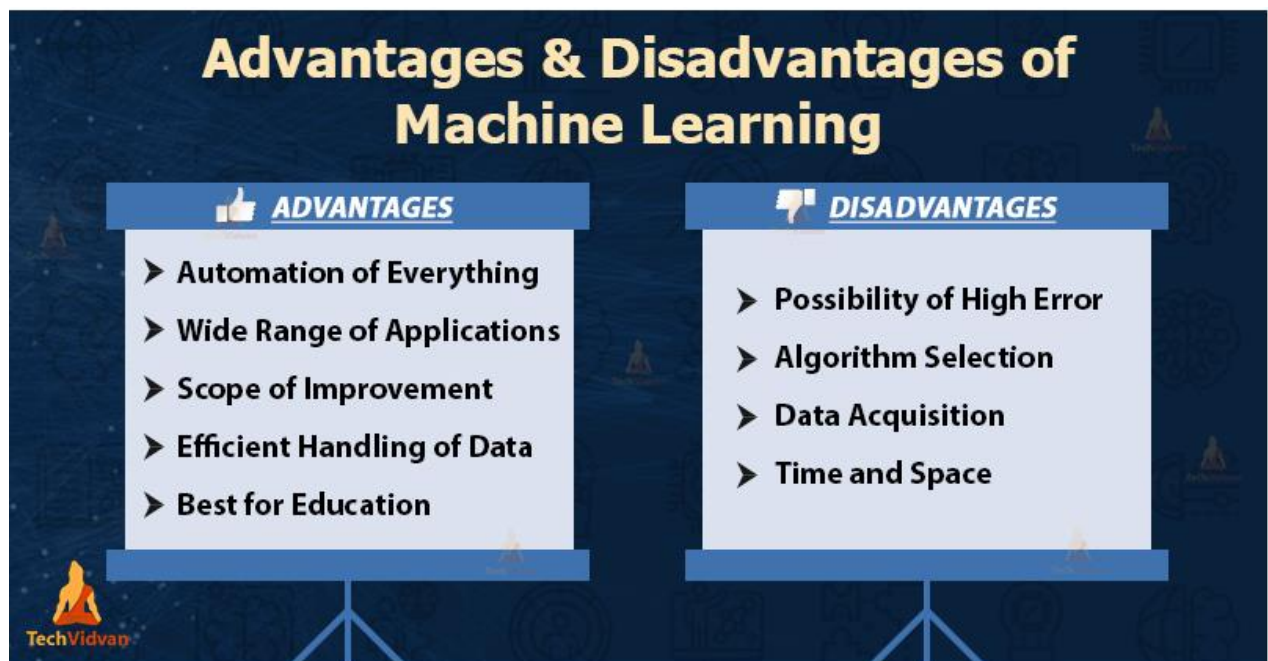
Classification : Un modèle de classification, une méthode d'apprentissage supervisé, tire une conclusion à partir des valeurs observées comme un ou plusieurs résultats sous une forme catégorique. Par exemple, le courrier électronique dispose de filtres comme la boîte de réception, les brouillons, le spam, etc. Le modèle de classification comporte un certain nombre d'algorithmes tels que la régression logistique, l'arbre de décision, la forêt aléatoire, la perception multicouche, etc. Dans ce modèle, nous classons nos données de manière spécifique et nous attribuons des étiquettes en conséquence à ces classes.

Clustering : Le clustering est une technique d'apprentissage automatique qui consiste à classer des points de données dans des groupes spécifiques. Si nous avons des objets ou des points de données, nous pouvons appliquer l'algorithme de mise en grappe pour les analyser et les regrouper en fonction de leurs propriétés et de leurs caractéristiques. Cette méthode de technique non supervisée est utilisée en raison de ses techniques statistiques. Les algorithmes de regroupement font des prédictions basées sur des données d'entraînement et créent des groupes sur la base de la similarité ou de la méconnaissance.

Scikit learn & Python



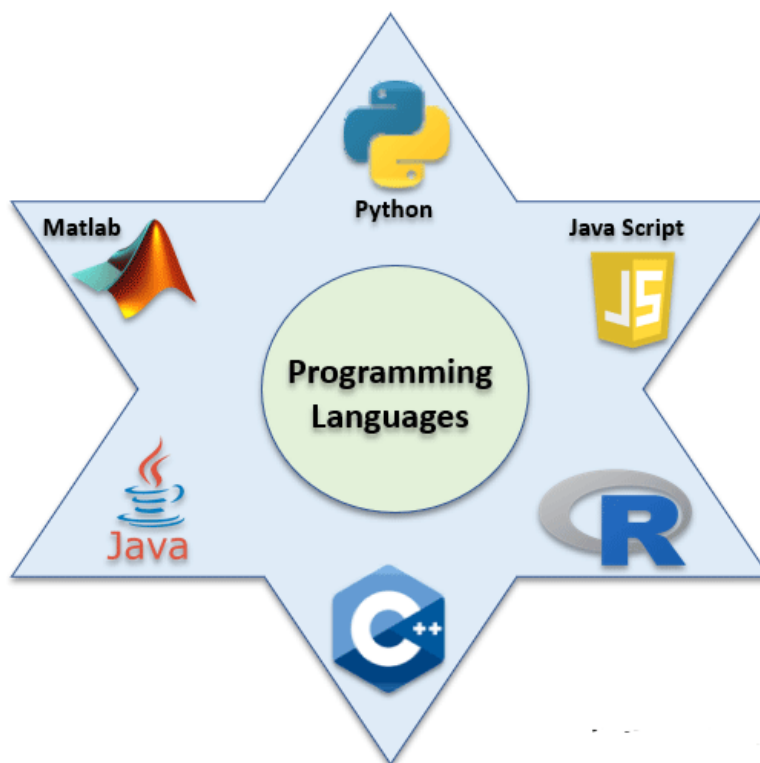
Avantages et inconvénients de l'apprentissage machine :



Langages de programmation pour l'apprentissage machine :

L'apprentissage machine est adopté par la plupart des entreprises pour construire et améliorer leurs produits et services. Pour mettre en œuvre ces algorithmes d'apprentissage machine, les entreprises ont le choix entre une grande variété de langages de programmation, à savoir python, C++, JavaScripts, Java, etc. car chacun de ces langages de programmation a ses propres caractéristiques, avantages et inconvénients. Il est bon d'en connaître plusieurs et il convient de choisir le langage de programmation spécifique à l'organisation en fonction du type d'application requis par celle-ci.

Jetons un coup d'œil sur les quelques langages de programmation populaires les plus utilisés pour l'apprentissage des machines :



Python

Définition :

Python est le langage de programmation open source le plus employé par les informaticiens. Ce langage s'est propulsé en tête de la gestion d'infrastructure, d'analyse de données ou dans le domaine du développement de logiciels. En effet, parmi ses qualités, Python permet notamment aux développeurs de se concentrer sur ce qu'ils font plutôt que sur la manière dont ils le font. Il a libéré les développeurs de contraintes de formes qui occupaient leur temps avec les langages plus anciens. Ainsi, développer du code avec Python est plus rapide qu'avec d'autres langages.

Il reste aussi accessible pour les débutants, à condition de lui consacrer un peu de temps pour la prise en main. De nombreux tutoriels sont d'ailleurs disponibles pour l'étudier sur des sites internet spécialisés ou sur des comptes YouTube. Sur les forums d'informatique, il est toujours possible de trouver des réponses à ses questions, puisque beaucoup de professionnels l'utilisent.

A quoi sert le langage Python ?

Les principales utilisations de Python par les développeurs sont :

- La programmation d'applications.
- La création de services web.
- La génération de code.
- La métaprogrammation.

Actuellement, la langue qui connaît la plus forte croissance au monde. Python a été conçu pour être lisible et facile à utiliser. Il utilise les concepts oops mais peut également être utilisé comme langage de script. Il est préférable pour le traitement du langage naturel et l'analyse sentimentale. Il possède presque tous les paquets nécessaires aux tâches d'apprentissage machine. Consultez le tableau ci-dessous pour voir quelques exemples de la

Task	Library
Compute Vision	OpenCV
Data pre-processing	Pandas
General-purpose machine learning	Scikit learn, TensorFlow, pytorch
data visualization	Matplotlib
Web application	django

bibliothèque d'apprentissage machine :

Nous avons également un ordinateur portable Jupyter, spécialement conçu pour la programmation python, qui permet de vérifier en temps réel la sortie de chaque ligne de code. Afin de développer votre propre algorithme à partir de zéro, vous pouvez utiliser la bibliothèque matrix max connue sous le nom de numpy et construire votre propre algorithme d'apprentissage machine. Python dispose d'une vaste communauté de forums ouverts comme stack overflow et GitHub. Google a récemment mis au point un excellent service de cloud computing gratuit, connu sous le nom de google Colab, que vous pouvez utiliser pour construire et former votre réseau en python à partir de zéro. Il est également fourni avec les GPU et les TPU, ce qui vous donne la liberté de coder dans un système à faible configuration.

Scikit-learn

Définition :

Sci-kit learns peut être considéré comme le cœur de l'apprentissage machine classique, qui est entièrement axé sur la modélisation des données au lieu de les charger, de les manipuler ou de les résumer. Il suffit de nommer une tâche et sci-kit learn peut l'exécuter efficacement. Une des bibliothèques les plus simples et les plus efficaces pour l'exploration et l'analyse de données, sci-kit learn est une bibliothèque open-source qui est construite sur NumPy, SciPy & Matplotlib. Elle a été développée dans le cadre du projet google summer code, qui est maintenant devenu une bibliothèque largement acceptée pour les tâches d'apprentissage automatique. Les apprentissages Sci-kit peuvent être utilisés pour préparer la classification, la régression, le regroupement, la réduction de la dimensionnalité, la sélection de modèles, l'extraction de caractéristiques, la normalisation et bien plus encore. L'un des inconvénients de l'apprentissage de la science et de la technologie est qu'il n'est pas pratique d'utiliser des données catégorielles.

Installation de scikit-learn :

Pour la plupart l'installation pip gestionnaire de paquets python peut installer python et toutes ses dépendances :

Cependant, pour les systèmes Linux, il est recommandé d'utiliser le **conda** paquets **conda** pour éviter les processus de génération possibles.

```
pip install scikit-learn
```

Pour vérifier que vous avez SCIKIT-LEARN, exécutez en Shell :

```
python -c 'import sklearn; print(sklearn.__version__)'
```

Application :

Le coronavirus (COVID-19) a récemment provoqué une préoccupation mondiale majeure.

Alors que le nombre de cas de coronavirus augmenterait, la propagation de **COVID-19** est une menace sérieuse pour la santé mondiale. Dans ce travail, nous allons essayer de prédire l'état de chaque pays à partir de la méthode régression logistique sous python à l'aide de bibliothèque scikit-learn.

Pour appliquer ces méthodes on doit importer la base de données des pays infectés par coronavirus : **covid_19_clean_complete.csv**

On doit d'abord importer les librairies concernées par :

```
#import des librairies l'environnement
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.model_selection import train_test_split
from matplotlib import cm
from sklearn.preprocessing import MinMaxScaler
from sklearn.linear_model import LogisticRegression

import plotly.express as px
import folium
```

Puis l'importation des données à partir de code python suivant :

```
#Importation des données

dt = pd.read_csv('C:/Users/MED-PC/Desktop/COVID-19-master/output.csv')
data = pd.read_csv('C:/Users/MED-PC/Desktop/COVID-19-master/input/corona-virus-report/covid_19_clean_complete.csv')

#spécifier le style de la table
dt.style.background_gradient(cmap="Blues", subset=['Confirmed', 'Active'])\
    .background_gradient(cmap="Greens", subset=['Recovered'])\
    .background_gradient(cmap="Reds", subset=['Deaths'])

# Les variables x et Y

feature_names = ['Confirmed', 'Active', 'Deaths', 'Recovered']
X = dt[feature_names]
y = dt['Classes']
```

Ça donne le tableau des données suivant :

Classes		Country/Region	Confirmed	Active	Deaths	Recovered
0	1	US	83836	81946	1209	681
1	1	China	81782	4310	3291	74181
2	1	Italy	80589	62013	8215	10361
3	1	Spain	57786	46406	4365	7015
4	1	Germany	43938	37998	267	5673
5	1	France	29551	22898	1698	4955
6	1	Iran	29406	16715	2234	10457
7	1	United Kingdom	11812	11082	580	150
8	1	Switzerland	11811	11489	191	131
9	1	South Korea	9241	4966	131	4144
10	1	Netherlands	7468	7027	435	6
11	1	Austria	6909	6748	49	112
12	1	Belgium	6235	5340	220	675
13	1	Canada	4042	4004	38	0
14	2	Turkey	3629	3528	75	26
15	2	Portugal	3544	3441	60	43
16	2	Norway	3369	3349	14	6
17	2	Brazil	2985	2902	77	6
18	2	Sweden	2840	2747	77	16
19	2	Australia	2810	2625	13	172
20	2	Israel	2693	2617	8	68
21	2	Malaysia	2031	1793	23	215
22	2	Denmark	2023	1932	41	50
23	2	Czechia	1925	1906	9	10
24	2	Ireland	1819	1795	19	5
25	2	Luxembourg	1453	1438	9	6
26	2	Ecuador	1403	1366	34	3

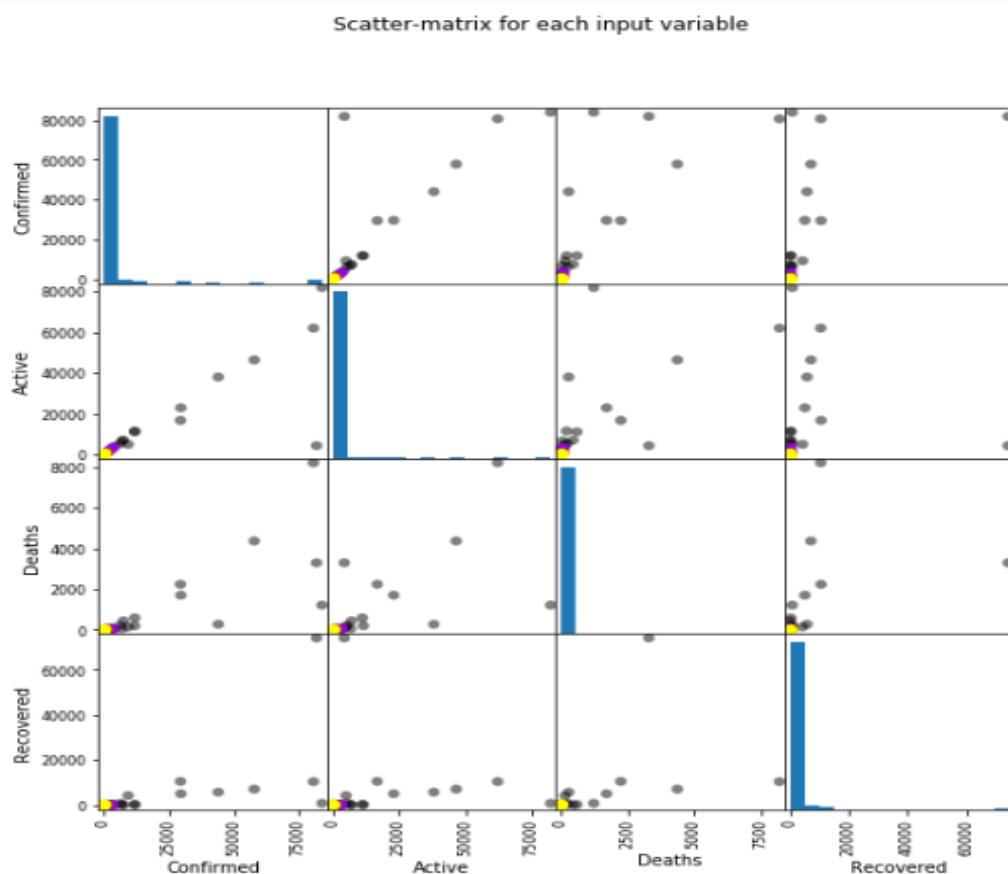
Une matrice de nuage de points est une matrice associée à n tableaux numériques (variables de données), X_1, X_2, \dots, X_n , de même longueur. La cellule (i, j) d'une telle matrice affiche le diagramme de dispersion de la variable X_i en fonction de X_j .

Nous montrons ici la fonction Plotly Express `px.scatter_matrix` pour tracer la matrice de diffusion pour les colonnes de la trame de données. Par défaut, toutes les colonnes sont prises en compte.

```
# scatter matrix

cmap = cm.get_cmap('gnuplot')
scatter = pd.plotting.scatter_matrix(X, c = y, marker = 'o', s=40, hist_kws={'bins':15}, figsize=(9,9), cmap = cmap)
plt.suptitle('Scatter-matrix for each input variable')
plt.savefig('COVID19_scatter_matrix')
```

Le résultat de ce code sous forme un graphe représentatif des quatre états :



Le découpage du jeu de données dans un projet de Machine Learning est une étape très importante qu'il ne faut pas négliger faute de quoi vous risquer de sur évaluer votre modèle (over-fitting) ou tout simplement le contraire (Under fitting). En effet par nature un modèle va coller (mais pas trop) à ses données d'entraînement.

Cette étape est donc une étape préalable mais aussi une étape d'optimisation qu'il ne faut pas laisser de côté.

```
#train et test des données

X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)
scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Regression Logistique

model = LogisticRegression(C=1.0, multi_class='auto', solver='liblinear')
model.fit(X_train, y_train)
print('Accuracy of Logistic regression classifier on training set: {:.2f}'
      .format(model.score(X_train, y_train)))
print('Accuracy of Logistic regression classifier on test set: {:.2f}'
      .format(model.score(X_test, y_test)))
```

La visualisation de données est un terme plus large qui décrit tout effort visant à aider les gens à comprendre l'importance des données en les plaçant dans un contexte visuel. Les modèles, les tendances et les corrélations peuvent être facilement montrés visuellement, ce qui pourrait autrement passer inaperçu dans les données textuelles. Créer des visualisations est assez facile, mais en créer de bonnes est beaucoup plus difficile. Il faut avoir le sens du détail et une bonne dose d'expertise pour créer des visualisations simples mais efficaces. Il existe aujourd'hui des outils de visualisation et des bibliothèques puissants qui ont redéfini le sens de la visualisation.

La beauté de l'utilisation de Python est qu'il offre des bibliothèques pour tous les besoins de visualisation de données. L'une de ces bibliothèques est **Folium**, qui est très pratique pour la visualisation de données géographiques (Géo data). La science des données géographiques (Géo data) est un sous-ensemble de la science des données qui traite des données de localisation, c'est-à-dire la description des objets et de leurs relations dans l'espace.

```
# map

m = folium.Map(location=[0, 0], tiles='cartodbpositron',
               min_zoom=1, max_zoom=4, zoom_start=1)

for i in range(0, len(data)):
    folium.Circle(
        location=[data.iloc[i]['Lat'], data.iloc[i]['Long']],
        color='crimson',
        tooltip = '<li><bold>Country : '+str(data.iloc[i]['Country/Region'])+
                  '<li><bold>Province : '+str(data.iloc[i]['Province/State'])+
                  '<li><bold>Confirmed : '+str(data.iloc[i]['Confirmed'])+
                  '<li><bold>Deaths : '+str(data.iloc[i]['Deaths']),
        radius=int(data.iloc[i]['Confirmed'])*1.1).add_to(m)

m
```

Nous obtiendrons une carte comme celle ci-dessous :



Interprétation :

- Chaque point représente les cas infectés par coronavirus dans tous les pays.
- Les points rouges sont des étiquettes pour chaque pays contient le nom de pays, le nombre de cas morts et le nombre des cas en rétablissement

Pour dessiner les graphes des qui représentent la somme des cas (Deaths and recoved)
et aussi les graphes qui représentent les régions par les morts et les confirmées :

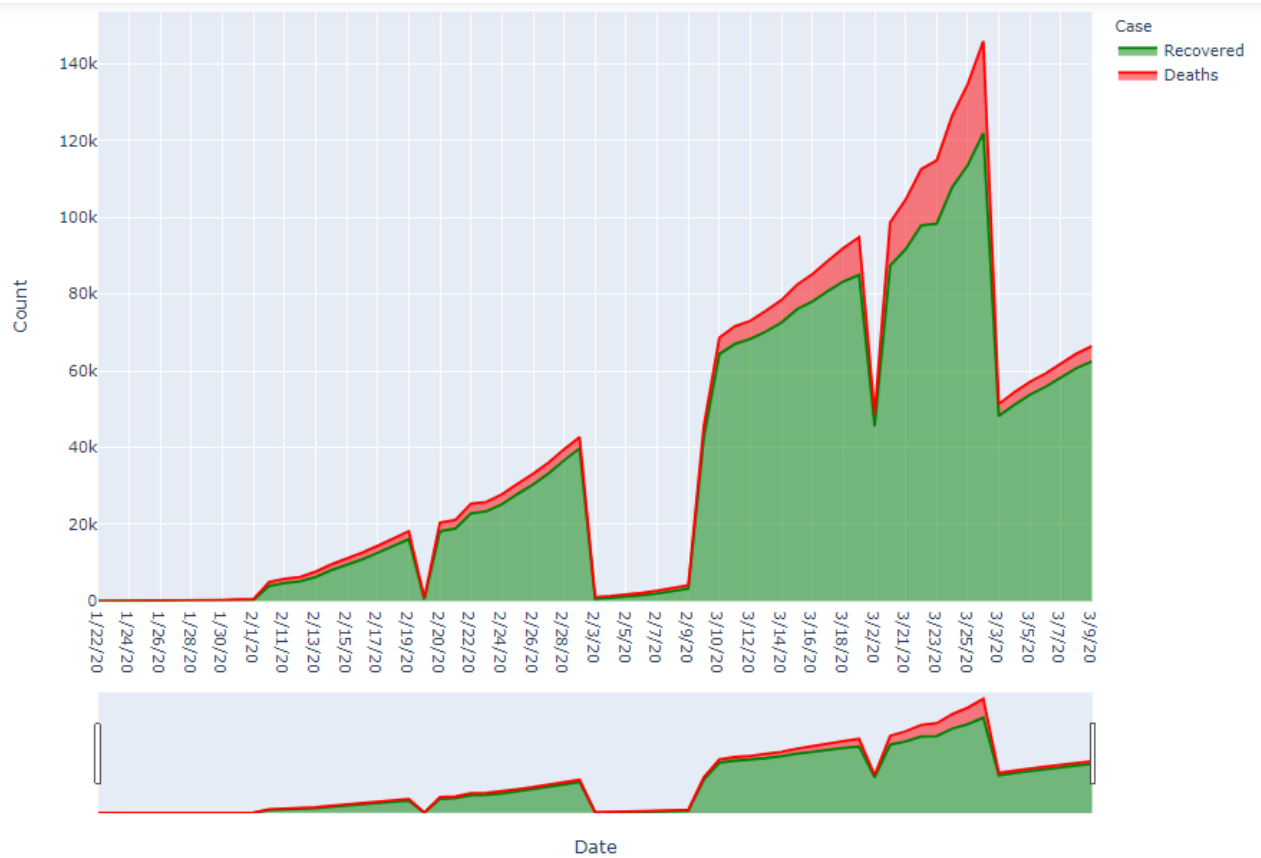
```
# graphe represente La somme des cas (Deaths and Recoved) par Date

temp = data.groupby('Date')['Recovered', 'Deaths'].sum().reset_index()
temp = temp.melt(id_vars="Date", value_vars=['Recovered', 'Deaths'],
                 var_name='Case', value_name='Count')
temp.head()

fig = px.area(temp, x="Date", y="Count", color='Case', height=800,
              title='Cases over time', color_discrete_sequence = ['green', 'red'])
fig.update_layout(xaxis_rangelslider_visible=True)
fig.show()

# graphe represente Les Region par Les (Deaths and Confirmed)

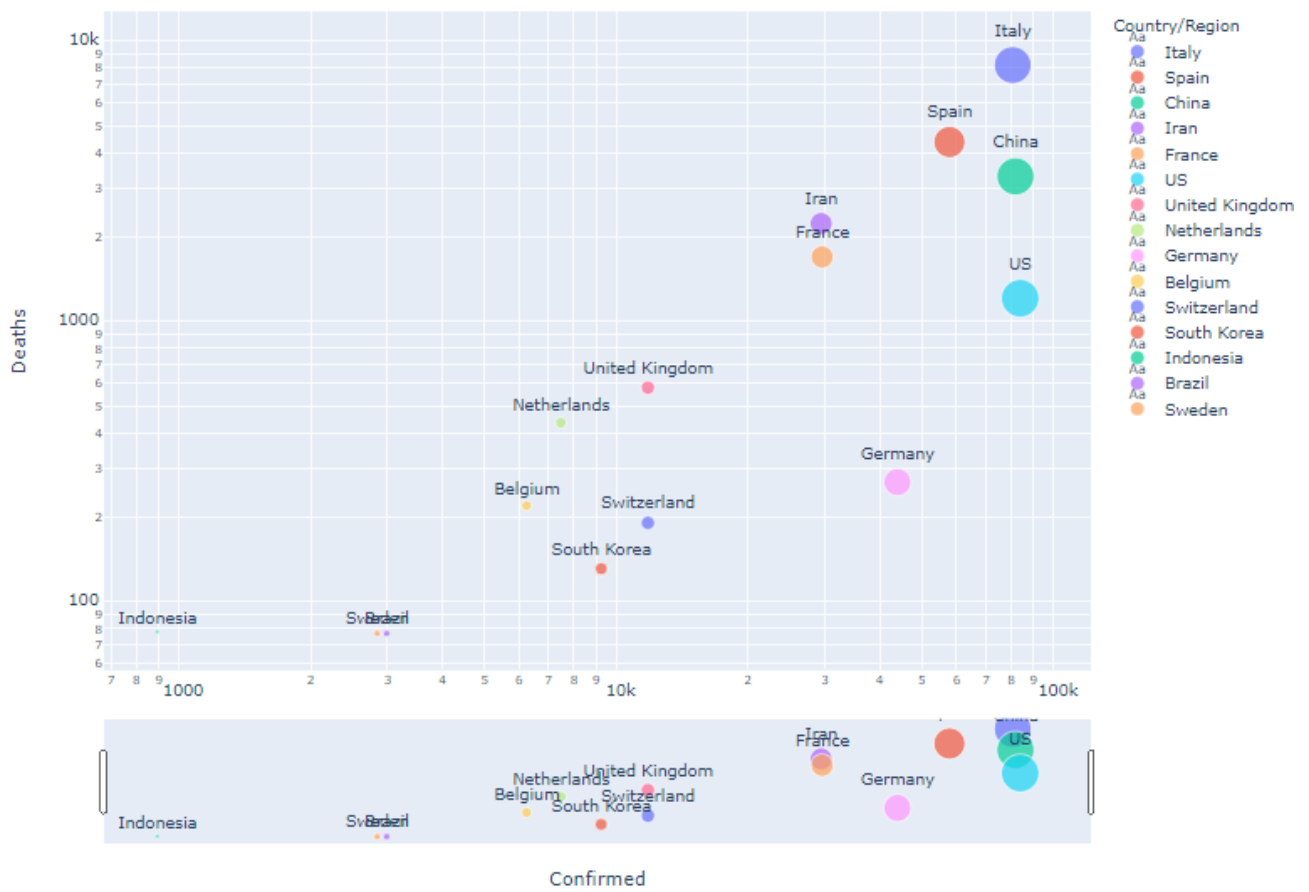
fig = px.scatter(dt.sort_values('Deaths', ascending=False).iloc[:15, :],
                 x='Confirmed', y='Deaths', color='Country/Region', size='Confirmed', height=800,
                 text='Country/Region', log_x=True, log_y=True, title='Deaths vs Confirmed')
fig.update_traces(textposition='top center')
fig.update_layout(xaxis_rangelslider_visible=True)
fig.show()
```



Interprétation :

- La courbe rouge représente développement du nombre total des cas morts au monde à partir du 01/22/2020 jusqu'à 03/08/2020
- La courbe verte représente développement du nombre total des cas recovred au monde à partir du 01/22/2020 jusqu'à 03/08/2020

Deaths vs Confirmed



Conclusion

Scikit-learn est largement écrit en Python, et utilise abondamment l'algorithme numpy pour des opérations d'algèbre linéaire et de tableaux à haute performance. En outre, certains algorithmes de base sont écrits en Cython pour améliorer les performances. Les machines à vecteurs de support sont implémentées par une enveloppe Cython autour de LIBSVM ; la régression logistique et les machines à vecteurs de support linéaires par une enveloppe similaire autour de LIBLINEAR. Dans de tels cas, l'extension de ces méthodes avec Python peut ne pas être possible.

Scikit-learn s'intègre bien avec de nombreuses autres bibliothèques Python, telles que matplotlib et plotly pour le traçage, numpy pour la vectorisation de tableaux, les cadres de données pandas, scipy, et bien d'autres.