



**INTRODUCTION À MACHINE LEARNING– Filière ASEDS INE1**

**Compte rendu du LAB 4 :**

Lab 4 : Simple and Multiple Linear Regression

**Réalisé par :**

Oussama louati

Achraf benomar

Khouloud taouchikht

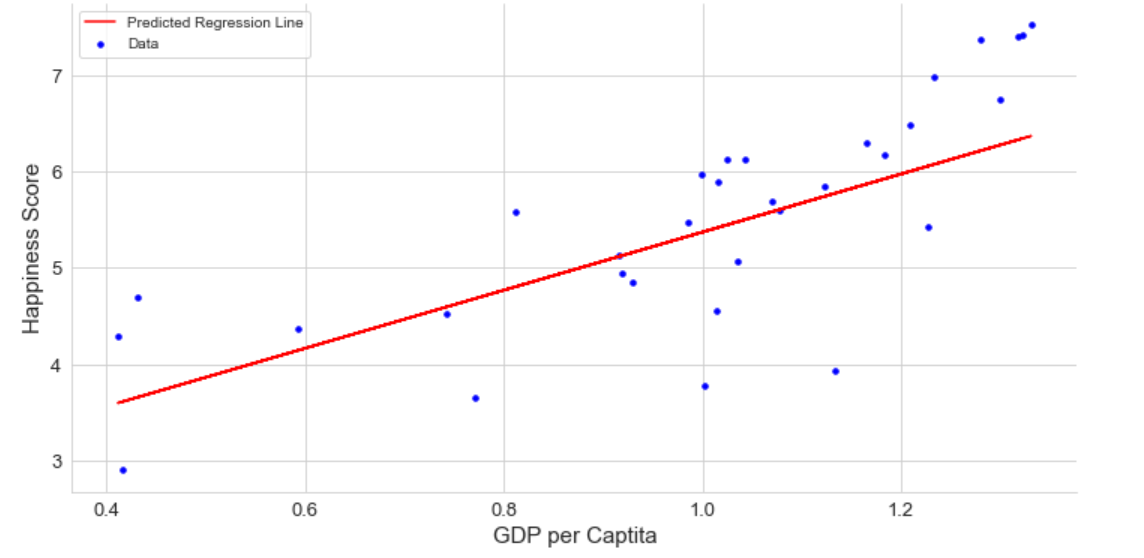
**Encadré par :**

Pr. Zineb El Akkaoui

**Année universitaire :**

2021/2022

1. **Interpret the obtained scatterplot :**



D’après le graphe,On constat que les deux variables « Happiness Score » et « GDP per capita » sont en relation quasiment linéaire.

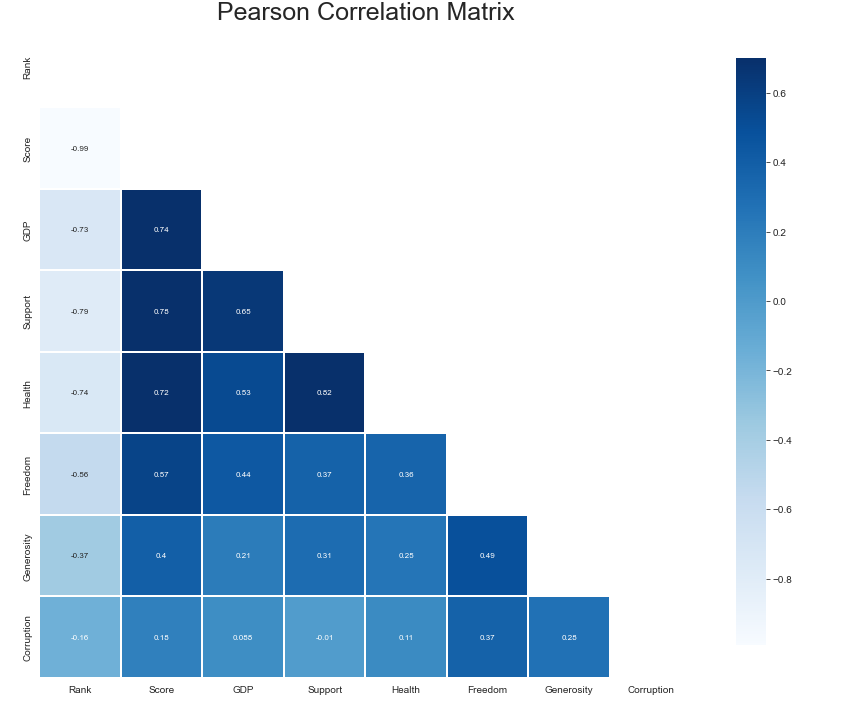
1. **Interpret the correlations :**

Le **coefficient de corrélation** permet de mesurer la force de la relation linéaire entre deux variables, c’est-à-dire lorsque le coefficient est proche de 1, la relation linéaire positive entre les variables est forte, et lorsque le coefficient est proche de −1, la relation linéaire négative entre les variables est forte.

D’après le matrice de corrélation de Pearson, On a pu remarquer qu’il y’ a des variables qui ont des relations linéaires fortes puisqui’ils ont des coefficients de corrélations de valeurs importantes, par exemple (GDP, Score), (Support, Score) et (Health, Support).

De meme ,il existe des variables avec des coefficients de corrélation de valeurs négatives, par exemple (Score, Rank), (Support, Rank) et (Health, Rank) ceci montre que les valeurs d'une de deux variable tendent à augmenter lorsque l'autre variable diminuent .

On trouve aussi des variables avec leurs coefficients de corrélation proche de zéro .D’ailleurs,les variables ne sont pas corrélées linéairement, par exemple (Support, Corruption), (Corruption, Rank) et (GDP, Corruption).

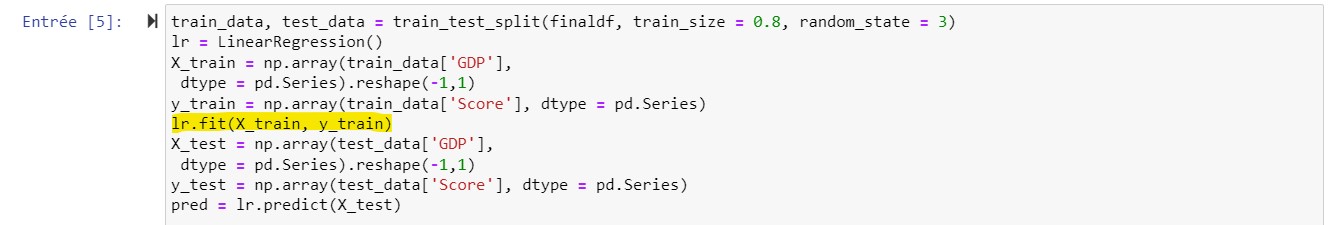


1. **Why is it advised to examine the scatterplot and the correlations among the features?**

Les scatterplots sont importants pour notre étude car ils peuvent montrer

. l'étendue de la corrélation, le cas échéant, entre les valeurs des quantités ou des phénomènes observés (appelés variables). S'il n'existe aucune corrélation entre les variables, les points apparaissent aléatoirement dispersés sur le plan de coordonnées.

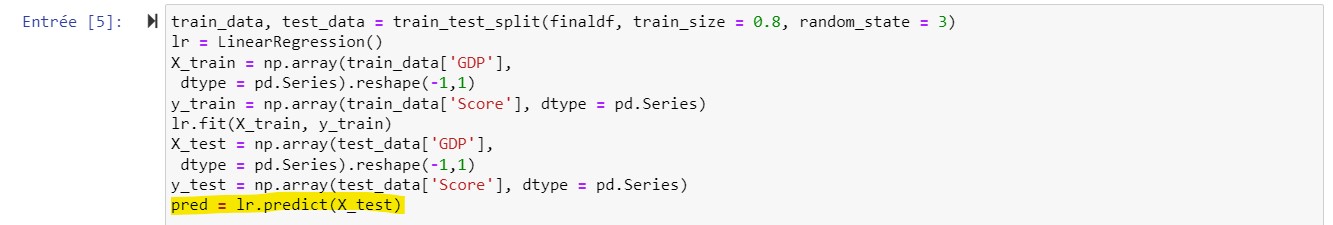
1. **How do we create the regression model? What are the features that has been used?**

****

Premierement on a employé la fonction LinearRegression() qui permet de construire un modèle en utilisant la régression linéaire des moindres carrés ordinaires.

ensuite,on a construit Notre modèle grace à la fonction lr.fit(X\_train, y\_train) qui prendre X\_train et y\_train comme parametres afin pour le modèle. le X\_train contient les variables d'entrée (GDP) alors que le y\_train contient la variable de réponse (Score).

1. **How do we apply the regression model?**

****

L’utlisation de la fonction lr.predict() en entrant X\_train comme parametre nous permet

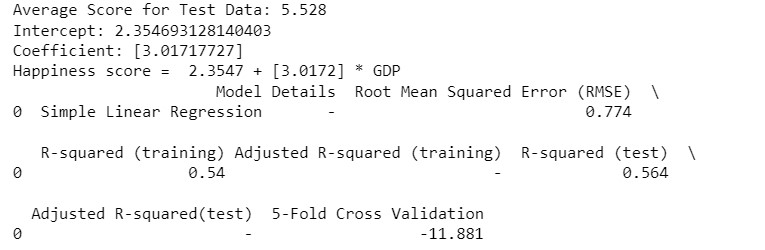
de prédire les étiquettes des valeurs des données sur la base du modèle formé.

1. **How the error term is computed?**



L’erreur quadratique moyenne (MSE) est calculée à l'aide de la fonction mean\_squared\_error() en utilisant « y\_train » et « pred » comme parametres

Le résultat après exécution du code est le suivant :



1. **Explain the difference between the two error term and the two adjusted R-squared**

La différence entre le R au carré et le R2  ajusté c’est que le R2 mesure la proportion de la variation de notre variable dépendante (Y) expliquée par nos variables indépendantes (X) pour un modèle de régression linéaire. Tandis que le R2 ajusté ajuste la statistique en fonction du nombre de variables indépendantes dans le modèle.

1. **Explain the 5-fold cross validation**

La validation croisée est une méthode statistique utilisée pour estimer la compétence des modèles de la machine learning

Il est couramment utilisé dans la machine learning appliqué pour comparer et sélectionner un modèle pour un problème de modélisation prédictive donné, car il est facile à comprendre, facile à mettre en œuvre et donne des estimations de compétences qui ont généralement un biais plus faible que les autres méthodes.

1. **What includes the pred variable?**

La variable pred applique le modèle de régression multivariée complexe\_model\_1 à l'ensemble de données de la manière suivante :





1. **Interpret the output results**

