**RECAPITULATION ET TEST D’EVALUATION**

1. Qu’est que la data science ?

*-> La data science: domaine qui utilise la statistique,l’algorithme informatique, les outils de l’intelligence artificielle pour analyser, exploiter des données,tirer des informations afin de prendre une décision*

1. Quels sont les principaux objectifs d’un projet de data science ?

*-> Transformer les données brutes en informations exploitables*

1. Quelle est la grande différence entre Data engineering, data analyste et data science ?

*->Data engineering: Prépare les données pour les data analyst et data scientist*

*-> Data analyste:analyse les données existantes*

*->Data science: modélisation de système pour faire la prédiction,prendre de décision*

1. Quels sont les types de données plus utilisées en data science ?

*->Données quantitatives(numériques)*

*->Données qualitatives(catégoriques)*

1. Quelles sont les grandes étapes d’un projet de data science ?

*->Collecte de données*

*->Analyse exploratoire des données*

*->Préparation des données*

*->Création de modèle*

1. Pourquoi la préparation des données est-elle essentielle ?

*-> La préparation des données est essentielle parce qu’il est important de travailler avec des données bien structurées,claires*

1. Qu’est-ce que l’analyse exploratoire des données (EDA) ?

*-> l’analyse exploratoire des données est une analyse utilisée pour comprendre les données et trier les informations pertinentes*

1. Qu’est-ce que l’évaluation d’un modèle ? Citez au moins deux métriques et donnez leur utilités.

*-> Évaluation d’un modèle: c’est une évaluation pour mesurer la performance,l’erreur d’un modèle*

*-Mean Absolute error: pour voir la moyenne des erreurs absolues*

*-Mean Square Error:pour voir la racine des erreurs absolues*

1. Qu’est-ce que la machine Learning ?

*-> C’est une domaine de l’intelligence artificielle qui consiste à apprendre automatiquement un système à partir des données pour faire de la prédiction*

1. Illustrez sous forme d’un tableau les trois grands types d’apprentissage.

|  |  |  |
| --- | --- | --- |
| Apprentissage supervisé | Apprentissage non supervisé | Apprentissage par reinforcement |

1. Illustrez à l’aide d’un tableau tous les algorithmes de machine Learning avec explication.

|  |  |
| --- | --- |
| Algorithmes | Explication |
| Régression linéaire | Algorithme pour prédire des valeurs continues |
| Régression logistique | Algorithme pour modéliser les données sous forme binaire |
| Random Forest |  |
| KNN |  |

1. Quelle est la différence entre surapprentissage et sous-apprentissage ?

*->Surapprentissage: apprentissage bien entraîné mais qui dépasse le norme*

*->Sous-apprentissage:apprentissage mal entraîné*

1. Comment choisir un algorithme pour un problème donné ?

*-> Choix d’un algorithme dépend de la forme du problème donné ou de la structure des données,l’objectif*

1. Pourquoi diviser les données en ensemble d’entrainement et de test ?

*->Parce qu’il faut entraîner le système avec des données d’entrainement et c’est à partir de cet entraînement qu’il est capable de tester avec d’autre données de test*

1. Quelle est la différence entre intelligence artificielle et data science ?

*->L’intelligence artificielle consiste à apprendre automatiquement un système à partir des données*

*-> Data science: consiste à collecter des données,les analysées,les interprétées*

1. Créer un projet personnel en data science bien explicite et pertinent.