**Rapport : Machine Learning Mini projet**

**Réaliser par : MQAM Youness et BELAKBIR Sara**

**En cadrent : M. HAJA Zakaria**

**Date : 16/06/2024**

**Rapport de Projet de Machine Learning : Prédiction des Performances des Élèves**

**. Introduction**

Ce rapport présente une analyse exploratoire des données concernant les performances des élèves dans une école. L'objectif principal est de développer un modèle capable de prédire les notes finales des élèves en fonction de diverses caractéristiques.

**2. Chargement et Exploration des Données**

Le dataset utilisé pour cette analyse contient des informations sur les performances académiques et les caractéristiques des élèves. Les données sont chargées à partir du fichier CSV "student-mat.csv".

**3. Distribution des Notes Finales**

Une visualisation de la distribution des notes finales des élèves est réalisée pour comprendre la répartition des performances dans l'échantillon.

**4. Matrice de Corrélation**

Une matrice de corrélation est calculée pour examiner les relations linéaires entre les variables numériques du dataset.

**5. Prétraitement des Données**

Les variables catégorielles sont encodées à l'aide de l'encodeur de labels, tandis que les variables continues sont normalisées en utilisant la mise à l'échelle standard.

**6. Modélisation**

Différents modèles de régression sont entraînés sur les données, notamment la régression linéaire, l'arbre de décision, la forêt aléatoire et le SVM (Support Vector Regressor).

**7. Évaluation des Modèles**

Les performances des modèles sont évaluées à l'aide de plusieurs métriques, notamment l'erreur absolue moyenne (MAE), l'erreur quadratique moyenne (MSE) et le coefficient de détermination (R2).

**8. Résultats**

Les résultats des modèles sont présentés sous forme de tableau, montrant les valeurs de MAE, MSE et R2 pour chaque modèle.

**9. Conclusion**

La régression linéaire est identifiée comme le modèle le plus performant pour prédire les notes finales des élèves, avec une MAE de 0.3268, une MSE de 0.2404 et un R2 Score de 0.7546.

Ce rapport met en évidence les principales étapes de l'analyse et de la modélisation des données sur les performances des élèves. Il offre des insights utiles pour comprendre les facteurs qui influent sur les résultats académiques des élèves et fournit un cadre pour le développement de modèles de prédiction précis.