**Dara PAK**

**Master 1 OIVM**

**TP ATDN**

Partie 1 : Régression avec Distribution Gaussienne sur des données générées

|  |  |
| --- | --- |
| Une image contenant diagramme, Tracé, ligne, capture d’écran  Description générée automatiquement | Une image contenant ligne, Tracé, diagramme, texte  Description générée automatiquement |
| Figure 1 : Generation des données de base | Figure 2 : Régression Lineaire et distribution gaussienne |

|  |  |
| --- | --- |
| Une image contenant Tracé, diagramme, ligne, texte  Description générée automatiquement | Une image contenant ligne, texte, Tracé, diagramme  Description générée automatiquement |
| Figure 3 : Distribution des résidus | Figure 4 Q-Q plot lineaire |

|  |  |
| --- | --- |
|  |  |
| Figure 5 : Régression Polynomiale | Figure 6 : Distribution des résidus Polynomiales |

|  |  |
| --- | --- |
|  |  |
| Figure 7 Q-Q Plot Polynomiales |  |

**Comparaison des Coefficients**

|  |  |
| --- | --- |
| R2 Lineaire : | 0.8406601636546879 |
| R2 Polynomiale : | -1.4225546966870883 |
| RMSE lineaire | 0.9962121504602562 |
| RMSE Polynomale : | 3.8844223714001873 |

Analyse R2 :

* Linéaire : Est de 84% de la variance, ce qui montre que le modèle est assez bien adapté pour capturer la relation entre les variables cibles et explicatives.
* Polynomial : Montre que le coefficient de détermination est surajoute, ce qui montre que le modèle polynomiale n’est pas très adapté pour capturer la relation entre les variables cibles et explicatives.

Analyse RMSE :

* Linéaire : 0.9962 montre que l'erreur moyenne entre les prédictions et les valeurs réelles est relativement faible pour le modèle linéaire, indiquant une bonne précision des prédictions.
* Polynomiale : 3.8844, qui est beaucoup plus élevé, indique que le modèle polynomial est moins précis par rapport au modèle linéaire.

**Comparaison des valeurs des coefficients de la régression Polynomiale avec et sans Ridge :**

|  |  |
| --- | --- |
| Lineaire | [[3.96846751]] |
| Polynomiale | [[ 0.          4.84100842 -0.45190593]] |

Partie II : Régression linéaire et polynomiale sur le jeu de données publiques ‘House Price’

|  |  |
| --- | --- |
|  |  |
| Figure 8 : Données variable cible et variable explicatives | Figure 9 : Regrission lineaire prediction |

|  |  |
| --- | --- |
|  |  |
| Figure 10 : Distributions residus Lineaire | Figure 11 : Q-Q plot residus Lineaire |

|  |  |
| --- | --- |
|  |  |
| Figure 12 : Régression polynomiale | Figure 13 : Comparaison régression polynomiale avec et sans Ridge |

**Comparaison des Coefficients**

|  |  |
| --- | --- |
| RMSE Linéaire | 56034.303865279944 |
| RMSE Polynomiale | 98483.61544430589 |
| RMSE Polynomiale avec Ridge | 79398.42392700305 |
|  |  |
| R^2 Linéaire | 0.5021486502718042 |
| R^2 Polynomiale | -0.5378702456449553 |
| R^2 Polynomiale avec Ridge | 0.0004247586899346345 |

Analyse des résultats :

Analyse R2 :

* Linéaire : Est de 50% de la variance, ce qui montre que le modèle est assez bien adapté pour capturer la relation entre les variables cibles et explicatives ce qui est bien meilleur que les 2 autres.
* Polynomial simple : La valeur négative de -0.5379 montre surajustement.
* Polynomial avec Ridge : Avec un R2 proche de zéro 0.0004, le modèle Ridge parvient à éviter le surajustement avec le modèle polynomial simple.

Analyse RMSE :

* Linéaire : Avec la valeur la plus faible parmi les trois modèles, le modèle linéaire a des prédictions relativement plus proches des valeurs réelles
* Polynomiale simple : Le RMSE élevé de 98483.62 montre une erreur importante dans les prédictions
* Polynomiale Ridge : Le RMSE diminue à 79398.42, ce qui est mieux que le modèle polynomial simple. La régularisation Ridge aide à réduire l'erreur, mais elle n'égale pas la simplicité et la précision du modèle linéaire.

**Comparaison des valeurs des coefficients de la régression Polynomiale avec et sans Ridge :**

|  |  |
| --- | --- |
| Régression Polynomiale | [ 0.00000000e+00  1.45549389e+02 -1.02504204e-02] |
| Régressions Polynomiale Ridge | [ 0.00000000e+00 -4.61750912e+00  4.10715169e-04] |

Partie III : Régression Logistique

Une image contenant texte, capture d’écran, diagramme, Tracé

Description générée automatiquement

Figure 14 : Distributions des prix des maisons

Analyse rapport de classification :

Une image contenant texte, capture d’écran, Police, logiciel

Description générée automatiquement

**Classe 0 (moins chère) :**

* Précision : 0.78, indique que parmi les maisons prédites comme "moins chères", 78% étaient correctes.
* Rappel : 1.00, signifie que le modèle a bien identifie toutes les maisons "moins chères".
* F1-score : 0.88, une très bonne moyenne entre la précision et le rappel.

**Classe 1 (plus chère) :**

* Précision : 1.00, ce qui indique que toutes les maisons prédites comme "plus chères" l’étaient effectivement.
* Rappel : 0.85, indiquant que le modèle a bien identifié 85% des maisons réellement "plus chères".
* F1-score : 0.92, qui montre une bonne contrepartie entre la précision et le rappel.

**Accuracy (Précision Globale) :** 0.90, montre que le modèle est correct dans 90% des cas.