

TP Machine Learning – Prédiction du prix des voitures

MASTER AI Année universitaire 2024-2025

Objectifs du TP

- Manipuler un jeu de données réel (prix de voitures)
- Mettre en œuvre un pipeline complet de machine learning supervisé
- Utiliser la régression linéaire pour la prédiction
- Évaluer un modèle et le déployer via une interface simple (Gradio)

1. Import des bibliothèques

Consigne et explication

| Question | Explication |
|--|--|
| Importez les bibliothèques nécessaires pour le traitement de données, la visualisation et le machine learning (<code>pandas</code> , <code>numpy</code> , <code>matplotlib</code> , <code>seaborn</code> , <code>scikit-learn</code>). | Les bibliothèques Python offrent des outils puissants pour la manipulation des données, leur visualisation et la création de modèles prédictifs. |

2. Chargement des données

Consigne et explication

| Question | Explication |
|---|--|
| Chargez le fichier <code>Data_Cars.csv</code> dans un DataFrame <code>pandas</code> et affichez les 5 premières lignes. | Cela permet de vérifier la structure et le contenu des données avant d'aller plus loin dans l'analyse. |

3. Prétraitement des données

Consigne et explication

| Question | Explication |
|--|--|
| <ol style="list-style-type: none">1. Renommez les colonnes pour plus de clarté2. Vérifiez la présence de valeurs manquantes3. Encodez la variable "Marque" en valeurs numériques | Un prétraitement rigoureux facilite l'analyse et améliore la qualité des modèles. L'encodage des variables catégorielles est indispensable pour le machine learning. |

4. Analyse exploratoire (EDA)

Consigne et explication

| Question | Explication |
|--|---|
| <ol style="list-style-type: none">1. Affichez les statistiques descriptives du DataFrame2. Visualisez la matrice de corrélation entre les variables | L'exploration des données aide à détecter les tendances, les anomalies, et à choisir les variables les plus pertinentes pour la modélisation. |

5. Suppression de colonnes peu corrélées

Consigne et explication

| Question | Explication |
|--|--|
| Supprimez la colonne "Chevaux" si elle n'est pas corrélée au prix. | Garder uniquement les variables pertinentes rend le modèle plus performant et plus simple à interpréter. |

6. Séparation des données

Consigne et explication

| Question | Explication |
|--|---|
| Séparez les variables explicatives (X) de la variable cible (y). | La séparation X/y permet d'entraîner le modèle sur les bonnes colonnes et d'évaluer ses performances. |

7. Création et entraînement du modèle

Consigne et explication

| Question | Explication |
|--|---|
| Créez et entraînez un modèle de régression linéaire. | La régression linéaire est un modèle prédictif simple permettant d'estimer la valeur d'une variable cible à partir de variables explicatives. |

8. Prédiction

Consigne et explication

| Question | Explication |
|---|--|
| Prédisez le prix des voitures sur l'ensemble de test avec votre modèle. | Cette étape permet de tester la capacité de généralisation du modèle sur des données non vues pendant l'apprentissage. |

9. Évaluation du modèle

Consigne et explication

| Question | Explication |
|--|--|
| Évaluez la performance de votre modèle avec l'erreur quadratique moyenne (MSE) et le score R^2 . | L'évaluation permet de quantifier l'écart entre les prédictions et la réalité, et d'ajuster le modèle si nécessaire. |

10. Bonus : Interface Gradio

Consigne et explication

| Question | Explication |
|--|---|
| Créez une interface simple avec Gradio pour permettre de prédire le prix d'une voiture à partir de ses caractéristiques. | Gradio permet de transformer rapidement un modèle en une application interactive accessible à tous. |

Remarques et conseils

- N'hésitez pas à documenter votre code par des commentaires.
- Sauvegardez votre notebook à chaque étape.
- Si vous êtes bloqué·e, posez des questions et travaillez en binôme.
- Pour aller plus loin, vous pouvez tester d'autres modèles ou visualisations !