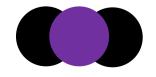
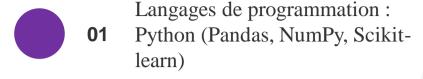
Portfolio – Soraya STOKKING (KAHLOUCHE) - Data Science



Introduction

Dotée d'une formation approfondie en data science, j'ai acquis un éventail de compétences me permettant d'aborder chaque projet avec une méthodologie rigoureuse. Mon parcours m'a permis de maîtriser des étapes essentielles, telles que le prétraitement des données, l'ingénierie des caractéristiques, ainsi que la mise en œuvre de modèles avancés en machine learning et deep learning. J'ai mené des projets diversifiés, tout en respectant les bonnes pratiques et en veillant à une communication claire et précise des résultats auprès des parties prenantes.





Machine Learning : Régression,
Classification, Clustering,
Réseaux de neurones

Deep Learning : CNN, RNN
avec TensorFlow et Keras

Visualisation de données : Matplotlib, Seaborn

Gestion des données : AWS, EMR, S3

06 Outils : Git, Jupyter Notebook, Docker

Compétences





- Machine Learning avec Python - Coursera
- Deep Learning Specialization Deeplearning.ai

Formation et Cours Complémentaires



1. Analyse des données des systèmes éducatifs

Contexte:

Analyse des données provenant de systèmes éducatifs de plusieurs pays pour identifier les opportunités de croissance et les tendances éducatives.

Étapes

Nettoyage des données, analyse exploratoire, et formulation de recommandations avec des visualisations graphiques pour le client.

Compétences clés

Python (Pandas, Matplotlib), Jupyter Notebook, statistiques descriptives.



2. Conception d'une application pour la santé publique

Contexte

Conception d'une application pour améliorer les habitudes alimentaires des utilisateurs dans le cadre de "Santé publique France".

Étapes

Analyse des besoins, préparation des données, prototypage de l'application, et présentation finale avec des visualisations et métriques.

Compétences clés

Python (Scikit-learn pour le prétraitement), visualisation avancée.





Contexte

Développement d'un modèle de prédiction pour optimiser la consommation électrique dans différents bâtiments.

Étapes

Préparation des données (extraction de caractéristiques, transformation des variables), modélisation de régression et séries temporelles, et évaluation du modèle.

Compétences clés

Modélisation de séries temporelles, régression, Scikit-learn, optimisation d'hyperparamètres.



4. Segmentation des clients pour un e-commerce

Contexte

Identification de segments de clients distincts pour personnaliser les offres et renforcer la fidélisation.

Étapes

Exploration et nettoyage des données client, clustering (K-Means, DBSCAN), et analyse des segments pour des actions marketing spécifiques.

Compétences clés

Clustering, K-Means, PCA pour réduction de dimension, visualisation des clusters.



5. Classification automatique de produits sur ecommerce

Contexte

Création d'un modèle pour classifier automatiquement les produits selon leur catégorie à partir des descriptions textuelles et des images.

Étapes

Traitement de données textuelles (vectorisation) et d'images (CNN), et entraînement de modèles pour la classification des produits.

Compétences clés

NLP, CNN, Scikit-learn, TensorFlow/Keras pour le deep learning.



6. Développement d'un modèle de scoring bancaire

Contexte

Création d'un modèle de scoring pour évaluer la solvabilité des clients dans le secteur bancaire.

Étapes

Nettoyage et transformation des données financières, création de modèles de classification (forêt aléatoire, SVM), et évaluation des performances.

DE GOGRING BABIGIE

Compétences clés

Modèles de classification, ROC AUC, visualisation et interprétation des résultats.



Contexte

Déploiement d'un modèle de machine learning dans un environnement cloud pour une utilisation à grande échelle.

Étapes

Préparation des données, parallélisation avec PySpark, et déploiement sur des plateformes cloud comme AWS et Google Cloud.

Compétences clés

PySpark, déploiement dans le cloud, mise en production de modèles, gestion des ressources cloud.



Ia méthode

Approche rigoureuse

- 1. Compréhension du problème
- 2. Exploration des données
- 3. Modélisation
- 4. Communication des résultats

Outils utilisés : Anaconda, Vs conde, Python, jupyter notebook



Ce qui me motive



Qu'est-ce qui inspire mon travail?

Au-delà de l'aspect technique, je suis passionnée par l'impact des données dans le monde actuel, aidant à transformer des informations en stratégies tangibles pour optimiser les services et anticiper les besoins.



Contactez-moi



Email

kahlouchesoraya27@gmail.com

Linkedin

linkedin.com/in/soraya-data-scientist



https://github.com/sorayadata?tab=repositories