

Thibault ROUSSEL

Data Scientist

PARCOURS

FORMATIONS & CERTIFICATIONS

Formations :

2020-2021 **Master Ingénieur Machine Learning – Data Scientist**
CentraleSupélec & OpenClassrooms

2017-2020 **Licence Mathématiques-Informatique**
UCO Angers

Certifications :

03/2022 **Implémentez vos bases de données relationnelles avec SQL**

04/2021 **Classez et segmentez des données visuelles**

03/2021 **Analysez vos données textuelles**

11/2020 **Décrivez et nettoyez votre jeu de données**

12/2019 **Initiez-vous à l'algèbre relationnelle avec le langage SQL**

02/2019 **Maîtrisez les fondamentaux d'Excel (version 2016)**

EXPERTISE

DOMAINES DE COMPÉTENCES

Langages **Python** (pandas, numpy, sklearn, tensorflow, nltk, seaborn, . . .) - Java

Base de données SGBD Oracle – SQL

Développement web Flask

Bureautique Excel

Collaboration Git, GitHub

Science des données Machine Learning, Deep Learning, Analyse de données

Mathématiques Algèbre, Analyse, Statistiques, Optimisation

Langues Anglais – Espagnol

EXPÉRIENCES

DÉTAIL DES PROJETS

— Recommandation de poste

Data Science

> Janvier 2022

Projets

Il s'agissait de recommander des emplois à des candidats en fonction des caractéristiques de chacun avec, en particulier, leur historique de candidatures ainsi que des données historiques des précédents postes occupés par les candidats. Il fallait correctement concaténer des tableaux Excel et définir le bon modèle et la bonne métrique d'évaluation afin d'automatiser la recommandation.

Environnement

Jupyter Notebook (python), Excel

— Dépistage du cancer du poumon basé sur l'Intelligence Artificielle

Data Science

> Janvier 2022

Projets

Il s'agissait d'attribuer un score de malignité selon la grille internationale LungRads à des nodules (cellules cancéreuses) pulmonaires. Nous avons étudié les nodules sous formes de scanners en plusieurs couches afin d'y faire une analyse pseudo 3D pour entraîner un réseau de neurone à déterminer la texture du nodule puis nous avons déterminé sa coupe centrale afin de donner à chacun un score de malignité.

Environnement

PyCharm, Jupyter Notebook (python), Excel

— Ingénieur Machine Learning – Data Scientist

CentraleSupélec et OpenClassrooms

> 2020-2021

Projets

- Programmer des algorithmes de Machine Learning.
- Entraîner des réseaux de neurones profonds.
- Modéliser des données complexes, comme des images et des textes.
- Mener un projet de recherche et réaliser une preuve de concept (POC).

- **Projet 1** : Conception d'une application au service de la santé publique.

Nettoyage de données, analyse et communication de résultats.

- **Projet 2** : Anticipation des besoins en consommation électrique de bâtiments.

Grâce à un modèle d'apprentissage supervisé dont on a adapté les hyperparamètres et transformé certaines variables pour améliorer les résultats du modèle.

- **Projet 3** : Segmentation des clients d'un site e-commerce.

Grâce à un modèle d'apprentissage non supervisé que l'on a également évalué et amélioré.

- **Projet 4** : Catégorisation automatique de questions.

Représentation graphique de données à grandes dimensions, prétraitement des données, extractions de features et réduction dimensionnelle.

- **Projet 5** : Classification d'images à l'aide d'algorithmes de Deep Learning.

Évaluation des performances d'un modèle, adaptation de ses paramètres, sélection d'un modèle d'apprentissage et transformation des variables pertinentes.

- **Projet 6** : Développement d'une preuve de concept.

En identifiant des sources d'informations fiables et pertinentes (veille) pour résoudre un problème de data-science.

- **Projet 7** : Participation à une compétition Kaggle !

Rédaction d'une note méthodologique afin de communiquer ma démarche de modélisation et utilisation d'un logiciel de version de code.

Environnement

PyCharm, Jupyter Notebook (python), Excel, PowerPoint, MySQL, Git, GitHub, Flask

— Mémoire projet

CHU d'Angers

> 2020

Projets

Planification du personnel du CHU d'Angers avec le logiciel Cplex d'IBM.

Aujourd'hui, la planification du personnel des CHU pose un réel problème. Il s'agissait alors, avec plusieurs doctorants, d'effectuer un travail d'optimisation poussé du personnel en fonction d'un grand nombre de contraintes générales et que celui-ci soit modifiable pour les contraintes propres au personnel.

Environnement

Cplex

— Télé-Enquêteur

THE LINKS - INSTITUT MOAÏ - NANTES

> Juin-Juillet 2018 - Mai-Juillet 2019 - Octobre 2020

Projets

Enquêtes de satisfaction auprès de clients de différentes banques, mutuelles, caisses de retraite etc. Il m'a également été demandé d'encoder des réponses et de nettoyer les données de sorties.

Environnement

Logiciels entreprise, Excel

— Stage en gestion de patrimoine

Marc Antoine Regnier – Tours

> Février 2019

Projets

Il s'agissait d'optimiser une multitude de feuilles Excel afin d'automatiser la réalisation d'audit patrimoniale pour Monsieur Regnier et ses 50 collaborateurs. J'ai suivi en amont une formation à l'utilisation poussée d'Excel.

Environnement

Excel