

Parcours formation diplômante Data Scientist – Titre RNCP niveau 7

De Mars 2021 à février 2022 – OpenClassrooms

Synthèse:

Parcours Data Scientist avec des cours réalisés en partenariat avec CentraleSupélec, sessions de mentorat et soutenance de 8 projets d'entreprise. Formation diplômante avec l'obtention d'un titre RNCP niveau 7 (bac+5) en Data Science.

Rôle:

Etudiant dans le cadre du parcours Data Scientist OpenClassrooms.

Réalisation:

- **Projet 1** (« Définissez votre stratégie d'apprentissage »).

- **Projet 2** (« Analysez des données de systèmes éducatifs ») :

Réalisation d'une analyse exploratoire (EDA) des données éducatives de la Banque Mondiale pour déterminer dans quels pays une start-up de formation en ligne peut se développer.

- **Projet 3** (« Concevez une application au service de la santé publique ») :

Réalisation d'analyses statistiques univariées et multivariées, du nettoyage des données (traitement des valeurs manquantes, des outliers) à partir des données Open Food Facts.

Analyse de la pertinence des variables du jeu de données sous la forme de représentations graphiques en vue de créer une application « Nutri-rein » pour évaluer la qualité des aliments (en fonction du Nutriscore) pour les insuffisants rénaux.

- **Projet 4** (« Anticipez les besoins en consommation électrique de bâtiments ») :

Réalisation d'une analyse exploratoire (EDA), feature engineering, mise en place d'un **modèle d'apprentissage supervisé** afin de prédire la consommation électrique et émissions de CO2 de la ville de Seattle, réglages des hyper-paramètres, évaluation des performances du modèle.

- **Projet 5** (« Segmentez les clients d'un site e-commerce ») :

Réalisation d'un preprocessing des données, mise en place d'un modèle d'apprentissage non supervisé (modèle de segmentation de clients pour orienter la communication du site), réglage des hyper-paramètres, évaluation des performances du modèle, mise en œuvre d'une simulation pour définir les délais de maintenance du modèle.

- **Projet 6** (« Classifiez automatiquement des biens de consommation ») :

Réalisation d'une étude de faisabilité d'un moteur de classification basé sur les descriptions et images d'objets pour l'automatisation de l'attribution de la catégorie de l'objet.

Pré-traitement des textes (stemming, lemming, vectorisation, approche NLP) et des images (redimensionnement des images, ajustement du contraste, extraction des features par Transfer Learning sur réseaux de neurones à convolution VGG16 / ResNet50), réduction de dimensions PCA, clustering k-means, visualisation TSNE.

- **Projet 7** (« Implémentez un modèle de scoring ») :

Construction d'un modèle de scoring afin de prédire la probabilité de défaillance d'un client pour l'accord ou refus d'un crédit pour une institution financière et d'un

dashboard interactif à destination des gestionnaires de la relation client permettant d'interpréter les prédictions du modèle et d'améliorer la connaissance client.

Préparation et analyse exploratoire des données à partir d'un kernel Kaggle existant, mise en place d'une métrique personnalisée pour évaluer le modèle (fonction de revenu net), ajustement du seuil de probabilité, rédaction d'une note méthodologique, déploiement du dashboard dans le cloud.

- **Projet 8** (« Déployez un modèle dans le cloud ») :

Développement d'une chaîne de traitements des données image (preprocessing, réduction de dimensions) dans le cadre d'une application mobile permettant de classer des images de fruit/légume suivant leur photo et de donner des informations sur le fruit/légume.

Mise en place de la chaîne de traitements à partir du framework ML pour les calculs distribués Apache Spark et de l'API PySpark et déploiement dans un architecture Big Data AWS (S3, Sagemaker).

Environnement: Anaconda, notebook Jupyter, Python (Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn), librairies Python spécialisées (Plotly, Yellowbricks, SHAP), Neuro computing (CNN VGG16 et ResNet50, Keras, TensorFlow), pattern recognition NLTK, architecture Big Data AWS (IAM, S3, EC2, Sagemaker), serveur d'applications Heroku, Framework ML (Apache Spark, Pyspark), Streamlit, Flask.

Ministère de la Défense SSF / Projet Sigle TMA MCO SI Marine nationale
Depuis Février 2020 à Octobre 2020 – Sopra Steria Group

Synthèse: Intégration dans l'équipe TMA MCO SI Marine nationale (maintenance corrective et évolutive autour de l'infocentre, appui aux équipes de développement).

Rôle: Référent fonctionnel et technique sur le domaine Infocentre.

Réalisation:

- MCO – maintenance corrective et évolutive (analyse des anomalies en production et recette, développement des correctifs)
- Réalisation d'assistances et de redressements de données.
- Participation aux chantiers Virtualisation du Marché Complémentaire: montée de version SAP BI, migration Oracle des jobs Talend.
- Développements des correctifs / évolutions sur les jobs Talend, les univers, les rapports Webi.
- Packaging des livraisons Infocentre.
- Préparation et animation de comités fonctionnels Infocentre.
- Préconisations techniques sur la supervision des processus d'exploitation Infocentre.

Environnement: Talend TOS 7.1.1, SAP BI 4.2 SP7 (Webi, UDT), Oracle 12/18c, HP ALM.

Ministère de la Défense DMAé / Projet Brasidas Refonte MCO aéronautique

Octobre 2019 à janvier 2020 – Sopra Steria Group

- Synthèse:** Refonte du MCO Aéronautique - Avant-vente sur les domaines Pilotage de la performance et Programmation financière et des activités des différents domaines (maintenance, logistique, achats, ...).
- Rôle:** Ingénieur avant-vente.
- Réalisation:**
- Participation aux réunions avec les éditeurs (SAP, Opus, Qlik) pour cadrer les attendus par rapport aux besoins du client.
 - Préparation des POC avec les éditeurs.
 - Réalisation de benchmarks sur les outils suivants (Suite SAP SAP BI + SAP Lumira + SAP BW/4Hana, SAP BPC, suite Opus, Qlik Sense) dans les domaines du Pilotage de la performance et de de la Programmation physico-financière (programmation dans les domaines financiers, des activités de maintenance, de logistique, des achats, ... par rapport aux exigences techniques et fonctionnelles du client).
 - Participation à la rédaction de l'offre et de la proposition commerciale.
- Environnement:** SAP BW/4Hana, SAP BPC 11, SAP BI 4 (Webi), SAP Lumira, Qlik Sense, Opus Suite (Opus10, Simlox, Catloc).

CETE Sud-Ouest / Analyse et conception projet infocentre RPLS

Octobre 2012 à mai 2013 – Sopra Steria Group

- Synthèse:** Analyse, conception, développement, recette fonctionnelles des évolutions 2012 et 2013 pour un infocentre national relatif au données recensant le parc locatif des bailleurs sociaux.
- Rôle:** Analyste-concepteur décisionnel.
- Réalisation:**
- Animation des ateliers.
 - Modélisation des évolutions au niveau des datamarts.
 - Mise à jour des spécifications d'alimentation de l'infocentre (partie ETL).
 - Mise à jour des documents de conception technique des univers BO et des restitutions Webi.
 - Modification des univers BO.
 - Rédaction des scénarios de tests en environnement outillé (HP ALM).
 - Exécution des scénarios de tests sous HP ALM.
- Environnement:** Talend TOS v4.3, BO XI R2 et XI 3.1, Designer BO, Webi, Oracle 11g R2, HP ALM (HP QC 11).

EDF TMA DPI / Appui fonctionnel / expertise BO sur projet BO Inventaire

Juin 2011 à septembre 2011 – Sopra Steria Group

- Synthèse:** Appui fonctionnel pour les équipes de développement et expertise BO sur le projet "BO Inventaire" (inventaire des déchets nucléaires).
- Rôle:** Expertise Business Objects.
- Réalisation:**
- Finalisation des spécifications fonctionnelles sur les évolutions du projet "BO Inventaire".
 - Développement des évolutions sur les rapports BO existants et de nouveaux rapports BO.
 - Conception et rédaction de scénarios de tests sous HP QC 10.
 - Exécution des tests sous HP QC 10.
- Environnement:** BO XI R3 SP2.1, HP QC 10, Oracle 10g.

CETE Sud-Ouest / Analyse et conception projet Infocentre RPLS

Février 2010 à juin 2011 – Sopra Steria Group

- Synthèse:** Analyse, conception et réalisation de l'infocentre relatif à l'application RPLS de recensement du parc des logements locatifs des bailleurs sociaux.
- Rôle:** Analyse-concepteur.
- Réalisation:**
- Analyse des besoins.
 - Animation des ateliers.
 - Modélisation du datawarehouse et des 2 datamarts (1 datamart statistiques et 1 datamart pour le contrôle opérationnel et les extractions de données).
 - Rédaction des documents d'analyse et des spécifications générales.
 - Rédaction des spécifications d'alimentation de l'infocentre (partie ETL).
 - Rédaction des documents de conception technique des univers BO et des restitutions Webi.
 - Participation au développement des traitements ETL.
 - Développement des rapports complémentaires BO pour les extractions réglementaires et restitutions personnalisées bailleur.
- Environnement:** Talend TOS v4.3, BO XI R2 et XI 3.1, Designer BO, Webi, Oracle 11g R2.

Rectorat de Bordeaux / Assistance MOE plateforme décisionnelle du PIAD

Décembre 2006 à Février 2010 – Sopra Steria Group

- Synthèse:** Assistance MOE relative à l'exploitation de la plateforme décisionnelle du PIAD, structure garante de l'architecture technique, de l'infrastructure, de l'intégration technique des développements académiques et nationaux.
- Rôle:** Assistance MOE.
- Réalisation:**
- Intégration technique des projets décisionnels nationaux réalisés par les équipes internes (en académies) et externes (prestataires).
 - Recette technique: contrôle des chaînes de traitement ETL, vérification de la conformité des univers et des datamarts, tests de performance des requêtes BO et optimisation.
 - Assistance à l'exploitation de la plateforme technique: mise en place d'environnements de recette, mise en production des applications décisionnelles nationales, administration (monitoring et tuning) de bases de données DB2, rédaction des documentations techniques.
 - Participation aux chantiers d'évolution de la plateforme technique: migration BO V6.5 vers BO XI R2, migration de DB2 UDB v8.2 vers DB2 V9, développement de KM spécifiques pour l'ETL Sunopsis/ODI, tests et benchs sur la data compression, les requêtes parallélisées et les bases fédérées DB2.
 - Assistance aux utilisateurs de niveau 1 et 2 pour l'analyse et la qualification des demandes d'assistance sur les problèmes liés à l'exploitation de l'outil ETL Sunopsis/ODI, sur les bases DB2 (monitoring) et sur BO (exploitation et gestion des serveurs Webi, monitoring).
 - Assistance aux équipes de développement académiques et nationales pour les projets décisionnels sur les aspects DB2, ETL, BO (tuning des requêtes BO, expertise sur le Designer BO, sur la conception d'univers et des rapports), interface avec les équipes support des éditeurs (IBM, Oracle, BO) et avec les équipes systèmes et réseaux du Rectorat de Bordeaux.
- Environnement:** Linux RedHat, Sunopsis V3, ODI V10.1.3.x, DB2 V9, BO V6.5 et XI R2, Tarentella Nagios, shell scripts Linux (bash, sed, awk, perl).

Neuf Telecom / Assistance MOE migration SI comptable Cegetel - Arcole

Février 2006 à Décembre 2006 – Sopra Steria Group

- Synthèse:** Assistance MOE sur la convergence des flux financiers et comptables entre les applications du SI de Cegetel et l'outil de gestion comptable Arcole de Neuf Telecom.
- Rôle:** Assistance MOE.
- Réalisation:**
- Rédaction de spécifications fonctionnelles et techniques.
 - Développement d'interfaces sous forme de chaînes de traitement ETL entre le SI Cegetel (Clarify, SAP,...) et l'outil de gestion comptable Arcole (Ares) de Neuf Telecom.
 - Intégration des flux financiers et du chiffre d'affaires de Cegetel dans le SI cible (Arcole).
 - Industrialisation des processus de production.
 - MCO (Maintenance des Conditions Opérationnelles) sur l'intégration de chaînes de

traitement ETL.

- Assistance technique et fonctionnelle des utilisateurs.
- Développement de rapports BO (tableaux de bord sur le suivi d'exploitation de l'intégration des flux financiers) à partir d'univers BO.

Environnement: Oracle 9i, ETL Sunopsis, UnixHP UX, shell scripts Unix, BO V6.5.