

Data Scientist training – RNCP level 7 certification

From March 2021 to February 2022 – OpenClassrooms in partnership with CentraleSupélec

Summary:

Data Scientist training with courses in partnership with CentraleSupélec, mentoring sessions and support for 8 corporate projects. Diploma training with the obtaining of a RNCP level 7 certification (bachelor's degree+5 years) in Data Science.

Role:

Student of the OpenClassrooms Data Scientist training.

Achievements:

- **Project 1** (« Define your learning strategy »).

- **Project 2** (« Analyse data from education systems ») :

Perform an exploratory data analysis (EDA) of the World Bank's educational data to determine in which countries an on-line training start-up may develop.

- **Project 3** (« Design an application for public health ») :

Realization of uni-variate and multivariate statistical analyses, data cleaning (processing missing values, outliers) from Open Food Facts data.

Analysis of the dataset variables relevance with graphical representations in order to create a "Nutri-kidney" application to evaluate food quality (based on Nutriscore) for patients with renal insufficiency.

- **Project 4** (« Anticipate the electrical consumption needs of buildings ») :

Completion of an exploratory data analysis (EDA), feature engineering, implementation of a supervised machine learning model to predict the electricity consumption and CO2 emissions of the city of Seattle, hyper-parameters settings, evaluation of model performance.

- **Project 5** (« Segmentation of customers of an e-commerce site ») :

Performing data preprocessing, setting up an unsupervised learning model (customer segmentation model to guide site communication), hyper-parameters setting, evaluating model performance, simulation to define model maintenance delay.

- **Project 6** (« Automatically classify consumer goods ») :

Completion of a feasibility study for a classification engine based on object text descriptions and object images in order to automate the assignment of object category.

Text and image preprocessing (stemming, lemming, vectorization, NLP approach for text, image resizing, contrast adjustment, features extraction by Transfer Learning on neural networks VGG16 / ResNet50 for images), PCA dimension reduction, k-means clustering, TSNE visualization.

- **Project 7** (« Implementation of a scoring model ») :

Building a scoring model to predict a customer's failure probability for financial institution loan approval / refusal from a bank and building an interactive dashboard for customer relationship managers to interpret predictions model and improve customer knowledge.

Data preparation and exploratory data analysis from an existing Kaggle kernel,

custom metric implementation to evaluate the model (net income function), probability threshold adjustment, methodological note writing, dashboard deployment in the cloud.

- **Project 8** (« Deploy a model in the cloud ») :

Image data processing chain development (preprocessing, dimension reduction) as part of a mobile application to classify fruit/vegetable images according to their photo and to give information about the fruit/vegetable.

Processing chain implementation from the Apache Spark ML framework (distributed calculations) and PySpark API and deployment in AWS Big Data architecture (S3, Sagemaker).

Tools used:

Anaconda, Jupyter notebook, Python (Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn), ML Python libraries (Plotly, Yellowbricks, SHAP), Neuro computing (CNN VGG16, ResNet50, Keras, TensorFlow), pattern recognition NLTK, AWS Big Data architecture (IAM, S3, EC2, Sagemaker), Heroku application server, ML framework (Apache Spark, Pyspark), Streamlit, Flask.

French Ministry of Armed Forces / SIGLE project for French Navy

From February 2020 to October 2020 – Sopra Steria Group

Summary:

Team integration for third party application maintenance of SIGLE project. Maintenance Repair and Operations (MRO) for French Navy information system (Corrective and evolutionary maintenance for BI systems, support to development teams).

Role:

Functional and technical leader in BI domain.

Achievements:

- MRO – corrective and evolutionary maintenance (fault analysis in test and production, patches development)
- Conduct assistance and data adjustments.
- Achievements for virtualization projects: SAP BI version upgrade, Talend jobs migration for Oracle version upgrade (Oracle 18c).
- Patches development / new developments for Talend jobs, SAP BI universes and Webi reports.
- Deliveries packaging.
- Preparation et facilitation of BI domain committees.
- Technical recommendations on the supervision of the BI systems operating processes.

Tools used:

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

French Ministry of Armed Forces / Brasidas project - Aeronautical MRO redesign

From October 2019 to January 2020 – Sopra Steria Group

- Summary:** Aeronautical MRO (Maintain, Repair and Operations) redesign. Pre-sales on performance management and financial programming for different domains (maintenance, logistics, purchasing, ...) of French air force information system.
- Role:** Pre-sales engineer in BI domain.
- Achievements:**
- Participation with meetings with software vendors (SAP, Opus, Qlik) to align the customer's expectations with the client's needs.
 - Proof of Concept (POC) preparation with software vendors.
 - Setting benchmarks on the following tools (SAP BI Suite + SAP Lumira + SAP BW/4Hana, SAP BPC, Opus tools suite, Qlik Sense) in Performance Management and Financial Programming domains (programming in the areas of finance, maintenance activities, logistics, purchasing, etc. in relation to the customer's technical and functional requirements).
 - Participation in the technical solution and business proposal drafting.
- Tools used:** SAP BW/4Hana, SAP BPC 11, SAP BI 4 (Webi), SAP Lumira, Qlik Sense, Opus Suite (Opus10, Simlox, Catloc).

CETE Sud-Ouest / RPLS BI project analysis and design

From October 2012 to May 2013 – Sopra Steria Group

- Summary:** Analysis, design, development, functional acceptance of the 2012 and 2013 software evolutions about BI application relating to data identifying housing stock for social housing rental companies.
- CETE Sud-Ouest is an entity of french Sustainable Development Office.
- Role:** BI Analyst.
- Achievements:**
- Workshop facilitation with product managers.
 - Modeling and development of data marts.
 - Update of ETL specifications for data warehouse and data marts loadings.
 - Update of technical design documents for BO universes and Webi reports.
 - Modifying BO universes.
 - Writing of acceptance tests with HP ALM.
 - Execution of test scenarios with HP ALM.
- Tools used:** Talend TOS v4.3, BO XI R2 et XI 3.1, BO Designer, Webi, Oracle 11g R2, HP ALM (HP QC 11).

EDF (french electricity company) / BO Inventaire project – BI expertise

From June 2011 to September 2011 – Sopra Steria Group

- Summary:** Functional support for BO (BI tool) development teams and expertise on the "BO Inventory" project (nuclear waste inventory).
- Role:** Business Objects expert.
- Achievements:**
- Participation in functional specifications on developments in the "BO Inventory" project.
 - IT development on existing BO reports and new BO reports.
 - Design and writing test scenarios under HP QC 10.
 - Running tests under HP QC 10.
- Tools used:** BO XI R3 SP2.1, HP QC 10, Oracle 10g.

CETE Sud-Ouest / RPLS BI project analysis and design

From February 2010 to June 2011 – Sopra Steria Group

- Summary:** Analysis, design and production of RPLS BI application for the census of the rental housing stock of social housing providers.
- Role:** BI analyst.
- Achievements:**
- Needs analysis.
 - Workshops facilitation.
 - Data warehouse and data marts modeling (1 statistical data mart and 1 data mart for operational control and data extractions).
 - Drafting of analysis documents and general specifications.
 - Drafting of ETL specifications.
 - Drafting of technical design documents for BO universes and Webi reports.
 - Participation in ETL developments.
 - Development of additional BO reports for regulatory extractions and customized reports for social housing agencies.
- Tools used:** Talend TOS v4.3, BO XI R2 et XI 3.1, Designer BO, Webi, Oracle 11g R2.

Rectorat de Bordeaux / Project manager support for PIAD BI platform

From December 2006 to February 2010 – Sopra Steria Group

- Summary:** Project management support for the operation of the PIAD BI platform, which guarantees the technical architecture, infrastructure and technical integration of academic and national developments.
- PIAD is an entity of educational authorities of Bordeaux (Rectorat de Bordeaux): this entity depends on french Ministry of Education.
- Role:** Project management support.
- Achievements:**
- BI national projects technical integration achieved by internal (academies) and external (service providers) teams.
 - Technical acceptance: control of ETL pipelines processing, verification of universes and data marts compliance, BO requests performance tests and optimization.
 - Support for the operation of the technical platform: setting up acceptance environments, production of national BI applications, DB2 databases administration (monitoring and tuning), drafting of technical documentation.
 - Participation in the technical platform development: BO V6.5 migration to BO XI R2, migration from DB2 UDB v8.2 to DB2 V9, specific KM (Knowledge Modules) development for the Sunopsis / ODI ETL, tests and bench marking on data compression, parallel queries and DB2 federated databases.
 - Level 1 and 2 user support for support requests analysis and qualification about problems related to the operation of the ETL Sunopsis / ODI tool, DB2 databases (monitoring) and BO (operation and management of Webi servers, monitoring).
 - Academic and national development teams support for BI projects on DB2, ETL, BO aspects (BO queries tuning, BO Designer, universe design and reports expertise), interface with software publisher teams (IBM, Oracle, BO) and with systems and networks teams of educational authorities of Bordeaux (Rectorat de Bordeaux).
- Tools used:** Linux RedHat, Sunopsis V3, ODI (Oracle Data Integrator) V10.1.3.x, DB2 V9, BO (Business Objects) V6.5 et XI R2, Tarentella Nagios, Linux shell scripts (bash, sed, awk, perl).

Neuf Telecom / Project management support – Accounting IT system migration

From February 2006 to December 2006 – Sopra Steria Group

- Summary:** Project management support on the convergence of financial and accounting flows between Cegetel's IT applications and the accounting management tool Arcole of Neuf Telecom.
- Neuf Telecom and Cegetel are phone companies. Cegetel merged with Neuf Telecom in 2008.
- Role:** Project Management support.
- Achievements:**
- Writing functional and technical specifications.
 - Development of interfaces in the form of ETL pipelines or processing chains between the Cegetel IS (Clarify, SAP,...) and the Arcole accounting management

tool (Ares) of Neuf Telecom.

- Integration of Cegetel's financial flows and revenue into the target Information System (Arcole).
- Industrialization of production processes.
- MRO (Maintenance, Repair and Operations) on ETL processing chains integration.
- Technical and functional support for users.
- BO reports (dashboards on operation monitoring of financial flows integration) from BO universe.

Tools used:

Oracle 9i, Sunopsis / ODI ETL, Unix HP-UX, shell scripts Unix, BO V6.5.