

Rapport personnel

Introduction

Ce rapport a pour objectif de synthétiser les compétences acquises et les apprentissages réalisés au cours de ma formation en Data Science et Machine Learning. Il s'appuie sur l'ensemble des projets menés, incluant un projet personnel technique et plusieurs projets de formation, afin de valoriser à la fois les compétences techniques développées, la capacité de prise de recul et les soft skills mobilisés.

Au-delà des aspects techniques, ce document met en évidence l'évolution de ma compréhension du métier de Data Scientist / Machine Learning, les progrès réalisés, les défis rencontrés ainsi que les axes d'amélioration identifiés. Il vise enfin à présenter un positionnement professionnel clair et cohérent, en lien avec les attentes du marché.

1.1 Présentation des projets

Projet personnel technique : Prédiction des ventes à J+7

Ce projet personnel a été mené de bout en bout, depuis l'analyse du besoin métier jusqu'à la mise en production d'une solution opérationnelle. Il consiste à prédire les ventes à J+7 à partir de données historiques et calendaires, sur un périmètre volontairement simplifié.

J'ai conçu l'ensemble du pipeline de prévision, incluant le feature engineering, la modélisation, l'évaluation des performances et la comparaison de plusieurs approches. La solution a ensuite été déployée via une API FastAPI et une application Streamlit, avec la mise en place de tests automatisés et d'une intégration continue.

Ce projet m'a permis de consolider ma compréhension du cycle complet Data Science / Machine Learning ainsi que des contraintes liées à la mise en production. Les principaux axes de progrès identifiés concernent le MLOps avancé, le monitoring, le feature engineering avancé et la vulgarisation technique.

Projets de formation

Analysez les données de systèmes éducatifs (30 h – projet validé)

Projet introductif axé sur l'analyse exploratoire de données, le nettoyage et la préparation des données, l'analyse statistique et la restitution de résultats à partir de visualisations.

Anticipez les besoins en consommation de bâtiments (30 h – projet validé)

Projet de régression supervisée visant à prédire des consommations énergétiques, avec un travail approfondi sur le feature engineering, l'évaluation de modèles et la prédiction de variables continues.

Classifiez automatiquement des informations (50 h – projet validé)

Projet de classification supervisée appliqué à des données textuelles, incluant le prétraitement des textes, l'entraînement de modèles de classification et l'analyse des performances à l'aide de métriques adaptées.

Déployez un modèle de Machine Learning (40 h – projet validé)

Projet centré sur la mise en production d'un modèle de Machine Learning, couvrant la sérialisation des modèles, la création d'API ML, les tests et la validation.

Initiez-vous au MLOps (Partie 1/2) (60 h – projet validé)

Projet orienté structuration de projets Machine Learning, versioning des modèles, mise en place de pipelines d'entraînement et bonnes pratiques MLOps.

Labellisez et appliquez des approches semi-supervisées en traitement d'images (40 h – projet validé)

Projet de vision par ordinateur mettant en œuvre des approches semi-supervisées, incluant l'extraction de features, le clustering et la labellisation d'images.

Confirmez vos compétences en MLOps (Partie 2/2) (50 h – projet validé)

Approfondissement des pratiques MLOps avec un focus sur le suivi d'expériences, le monitoring des modèles, la gestion du cycle de vie ML et l'industrialisation.

Concevez et déployez un système RAG (60 h – projet validé)

Projet consacré à la conception d'un système Retrieval-Augmented Generation, intégrant vectorisation, recherche sémantique, LLM et architecture IA moderne.

Évaluez les performances d'un LLM (60 h – projet validé)

Projet dédié à l'évaluation de modèles de langage, combinant métriques spécifiques aux LLM, analyses qualitatives et quantitatives, et esprit critique sur les résultats.

Réalisez le cadrage d'un projet IA (40 h – projet validé)

Projet de cadrage stratégique portant sur l'analyse des besoins métiers, la définition de cas d'usage et la gestion de projet IA.

Concevez un système de recommandations pour une agriculture optimisée par les données (80 h – projet validé)

Projet de recommandation orienté aide à la décision, combinant analyse de données métier, modélisation avancée et compréhension des enjeux sectoriels.

1.2 Compétences techniques acquises

Les projets réalisés au cours de la formation, ainsi que le projet personnel technique, m'ont permis de développer un ensemble cohérent de compétences couvrant l'ensemble du cycle Data Science / Machine Learning, depuis l'analyse des données jusqu'à la mise en production des modèles.

Les compétences acquises sont présentées ci-dessous par grands domaines, avec une auto-évaluation du niveau atteint.

Programmation & environnement de travail — 7/10

- Python (programmation orientée data)
 - SQL (requêtes, jointures, agrégations)
 - Gestion d'environnements (venv, Poetry)
 - Organisation et structuration de projets Python
-

Analyse de données & statistiques — 9/10

- Analyse exploratoire de données (qualitative et quantitative)
 - Statistiques descriptives
 - Analyse de corrélations
 - Visualisation de données (graphiques, tendances, distributions)
 - Interprétation critique des résultats
-

Préparation des données — 8/10

- Nettoyage de données
- Feature engineering
- Gestion des valeurs manquantes
- Agrégation de sources multiples

- Prévention de la fuite de données
-

Machine Learning (tabulaire & séries temporelles) — 7/10

- Entraînement, test et évaluation de modèles ML
 - Régression et classification
 - Optimisation de modèles
 - Baselines et comparaison de modèles
 - Prédiction
 - Séries temporelles à horizon court
 - Feature importance
 - SHAP values
-

Deep Learning & IA avancée — 6/10

- Utilisation de modèles de Deep Learning
 - Compréhension des architectures (CNN, embeddings)
 - Identification des limites et conditions d'utilisation
-

NLP, LLM & RAG — 6/10

- Traitement du langage naturel (NLP)
- Vectorisation et embeddings
- Vector stores
- Conception de systèmes RAG

- Intégration de LLM
 - Conception de chatbots
 - Évaluation des performances de modèles de langage
-

MLOps & industrialisation — 6/10

- Principes MLOps
 - Suivi d'expériences avec MLflow
 - Création de pipelines Machine Learning
 - Sérialisation et versioning de modèles
 - Monitoring des modèles
 - Profiling des performances
 - Mise en production de modèles ML
-

Déploiement & services ML — 8/10

- Création d'API REST avec FastAPI
 - Validation métier des entrées
 - Exposition d'APIs publiques
 - Déploiement cloud (Render)
 - Développement d'applications Streamlit
 - Déploiement via Streamlit Community Cloud
-

DevOps, CI/CD & conteneurisation — 7/10

- Git et GitHub
 - Intégration continue (GitHub Actions)
 - Tests automatisés
 - Tests unitaires et couverture de code
 - Docker et DockerHub
-

Bases de données — 8/10

- Gestion de bases de données relationnelles
 - PostgreSQL
 - Connexion entre applications et bases SQL
-

Qualité logicielle & robustesse — 8/10

- Tests unitaires avec pytest
 - Tests d'API
 - Validation des entrées
 - Gestion des erreurs
 - Robustesse des applications en production
-

Cadrage, communication & métier — 7/10

- Cadrage professionnel de projets Data / ML
- Analyse des besoins métiers
- Traduction d'un besoin métier en solution data

- Réalisation de Proof of Concept (PoC)
 - Interprétation métier des résultats
 - Rédaction de rapports synthétiques
 - Création de supports de restitution clairs
 - Vulgarisation technique
-

1.3 Évaluation et synthèse des compétences

Cette auto-évaluation met en évidence des **points forts marqués** en analyse de données, préparation des données, déploiement de solutions ML et qualité logicielle. Les compétences en Machine Learning, MLOps et IA avancée sont solides et en cours de consolidation, avec des axes de progression clairement identifiés.

L'ensemble de ces compétences a été mobilisé et validé à travers des livrables concrets : notebooks d'analyse, modèles entraînés, APIs, applications Streamlit, pipelines ML, tests automatisés, rapports et démonstrations en ligne.

2.1 Axes d'amélioration identifiés

Avec le recul, plusieurs axes d'amélioration se dégagent clairement.

Sur le plan technique, je souhaite approfondir en priorité la **gestion du code en Deep Learning et en NLP**, notamment autour des **LLM** et des **architectures RAG**, afin de gagner en maîtrise et en autonomie sur ces approches récentes.

Je souhaite également développer l'usage de **formats et outils d'optimisation**, tels que **ONNX**, pour améliorer les performances et la portabilité des modèles. Un autre axe majeur concerne la **justification mathématique** des modèles de Machine Learning, de Deep Learning et des LLM, dans le but de mieux comprendre leurs mécanismes internes et de renforcer la pertinence des choix techniques.

Par ailleurs, je souhaite approfondir les **pratiques avancées de monitoring en production**, ainsi que le **déploiement cloud de solutions ML plus complexes**. Une meilleure compréhension des **réseaux de neurones convolutionnels (CNN)** et du **feature engineering avancé** constitue également un axe de progression identifié.

Enfin, un effort particulier sera porté sur la **vulgarisation technique**, afin de mieux expliquer des concepts complexes à des publics non techniques et de faciliter la prise de décision métier.

Ces axes traduisent une volonté de renforcer la profondeur technique tout en améliorant la lisibilité et l'exploitabilité des solutions développées.

2.2 Défis à relever

Les principaux défis identifiés concernent le **renforcement des compétences en Machine Learning appliqué et en MLOps**, en particulier la **maîtrise avancée du code autour des LLM et des systèmes RAG**.

L'**industrialisation des modèles** via des pratiques MLOps solides constitue également un défi central, incluant le **monitoring des performances**, le suivi des modèles en production et l'anticipation de leurs limites dans des environnements réels.

Un autre défi majeur réside dans la capacité à **justifier mathématiquement les modèles** tout en étant capable de les **vulgariser efficacement**, afin de rendre les choix techniques compréhensibles pour des parties prenantes non spécialisées.

Enfin, la poursuite de **projets concrets en Data Science et Machine Learning** représente un enjeu clé pour renforcer simultanément :

- la compréhension des problématiques métiers,
- la maîtrise technique,
- et la capacité à proposer des solutions réellement adaptées aux besoins des organisations.

2.3 Évolution méthodologique – Ce que je ferais différemment

Si je devais recommencer certains projets avec davantage de temps, j'adopterais une méthodologie légèrement différente. Je consacrerai davantage de temps à la **recherche technique approfondie**, en m'appuyant sur la **documentation officielle**, des **ressources pédagogiques spécialisées** et des **échanges avec des communautés techniques**, afin d'approfondir ma compréhension des outils et des concepts utilisés.

Cette approche viserait à **renforcer ma capacité à concevoir, structurer et écrire du code de manière autonome**, tout en consolidant les **fondamentaux techniques**. L'objectif serait de mieux maîtriser les mécanismes sous-jacents des solutions développées et de gagner en efficacité et en robustesse sur le long terme.

2.4 Apports de la formation et des expériences

La formation et les projets réalisés m'ont permis de **structurer une démarche Data Science de bout en bout**, depuis l'analyse du besoin métier jusqu'à la mise en production de solutions exploitables. Ils m'ont également apporté une **meilleure compréhension des contraintes liées à la mise en production**, notamment en termes de robustesse, de maintenance et de suivi des modèles.

Ces expériences m'ont permis d'acquérir une **vision plus réaliste des enjeux techniques et métiers**, et ont renforcé ma capacité à **prendre du recul sur les choix techniques**, à en identifier les limites et à proposer des solutions adaptées au contexte et aux objectifs des organisations.

2.5 Évolution de ma perception du rôle de Data Scientist / Machine Learning

Avant cette formation, ma perception du métier de Data Scientist était principalement orientée vers l'**analyse de données**, des **statistiques légères** et la **création de tableaux de bord**.

Aujourd'hui, grâce aux nombreux projets réalisés, j'ai acquis une compréhension beaucoup plus **concrète et complète** de ce rôle. Je perçois désormais le Data Scientist et le Machine Learning Engineer comme des métiers **larges et diversifiés**, nécessitant une forte **expertise technique** et une **interaction étroite avec les enjeux métiers**.

Ces rôles sont également **évolutifs**, s'adaptant en permanence aux avancées technologiques rapides et à la diversité des contextes d'application. Cette richesse et cette diversité ouvrent de nombreuses opportunités et rendent le métier particulièrement stimulant et exigeant.

3. Mise en valeur des soft skills pour le métier de Data Scientist / Machine Learning

Les projets réalisés tout au long de la formation, ainsi que le projet personnel technique, m'ont permis de développer et de renforcer des **soft skills essentiels** au métier de Data Scientist / Machine Learning, en complément des compétences techniques acquises.

L'**esprit d'analyse et de résolution de problèmes** a été mobilisé de manière constante, notamment pour structurer des problématiques complexes, identifier les variables pertinentes et proposer des solutions adaptées aux contraintes techniques et métiers.

La **curiosité intellectuelle et l'apprentissage continu** ont également été au cœur de ma démarche, à travers l'exploration régulière de nouvelles approches, outils et technologies, qu'il s'agisse de Machine Learning, de MLOps ou d'IA générative.

La **communication et la vulgarisation** ont constitué un axe important, avec la capacité à expliquer des concepts techniques à des publics variés, à restituer les résultats de manière claire et à faciliter la compréhension et la prise de décision.

Les projets m'ont aussi permis de développer le **travail en équipe et la collaboration transverse**, en intégrant les enjeux métiers et techniques dans un cadre collaboratif, et en tenant compte des attentes des différentes parties prenantes.

L'**adaptabilité et l'agilité** se sont renforcées face à des contextes techniques variés et en évolution, nécessitant des ajustements réguliers de méthodes et de solutions.

La **gestion de projet et l'autonomie** ont été mobilisées à travers l'organisation du travail, la priorisation des tâches et la conduite de projets de bout en bout.

Enfin, la **créativité et la réflexion hors des sentiers battus** ont été sollicitées pour explorer des solutions alternatives et innovantes, notamment dans le choix des modèles, des architectures et des outils.

Par ailleurs, plusieurs soft skills transverses se sont consolidées au fil des projets, notamment l'**esprit critique**, avec la capacité à remettre en question les résultats et les choix techniques, la **rigueur**, traduite par une attention particulière portée à la qualité du code, des données et des livrables, ainsi que le **sens des responsabilités**, impliquant une prise de conscience des impacts métiers et techniques des décisions prises.

Conclusion

Cette formation en Data Science et Machine Learning m'a permis d'acquérir un **socle de compétences solide et cohérent**, couvrant l'ensemble du cycle de vie d'un projet data : analyse et préparation des données, modélisation, évaluation, déploiement et suivi en production. Les projets réalisés ont renforcé mes compétences en analyse de données, Machine Learning, déploiement de solutions ML, MLOps et qualité logicielle, tout en développant une capacité à relier les choix techniques aux enjeux métiers.

Au-delà des compétences techniques, les apports globaux de la formation résident dans la **structuration d'une démarche professionnelle**, la compréhension des contraintes réelles de mise en production et le développement d'une posture critique face aux résultats et aux solutions proposées. Cette formation m'a également permis d'acquérir une vision réaliste et opérationnelle du métier de Data Scientist / Machine Learning, fondée sur des projets concrets et des cas d'usage variés.

À l'issue de ce parcours, mon **positionnement professionnel** s'inscrit clairement dans un rôle de **Data Scientist orienté Machine Learning et mise en production**, capable de concevoir des solutions exploitables, robustes et adaptées aux besoins des organisations, tout en assurant leur déploiement et leur suivi.

Les **perspectives d'évolution** identifiées portent principalement sur l'approfondissement des pratiques MLOps avancées, le monitoring en production, le feature engineering avancé, les architectures Deep Learning et LLM/RAG, ainsi que la justification mathématique et la vulgarisation technique des modèles. La poursuite de projets concrets en Data Science et Machine Learning constitue une prochaine étape essentielle pour consolider ces compétences et accompagner une montée en expertise progressive.