

Rapport de conduite de projet Data & ML

Prévision des ventes à J+7 – Projet personnel professionnel

1. Contexte et analyse des besoins

1.1 Présentation (organisation et contexte)

Le projet s'inscrit dans un **contexte retail de grande distribution**, inspiré du dataset public *M5 Forecasting*.

L'organisation considérée est une **entreprise de grande taille**, opérant sur plusieurs territoires, disposant d'un volume important de données de ventes quotidiennes.

Enjeux stratégiques liés à la donnée :

- anticiper les ventes à court terme afin d'améliorer la planification,
- réduire l'incertitude dans la prise de décision commerciale,
- mieux exploiter les données historiques existantes.

Maturité data & ML :

- Données structurées et historisées disponibles.
 - Utilisation principalement descriptive de la donnée (reporting).
 - Faible automatisation de la prévision via des modèles prédictifs.
-

1.2 Collecte et analyse du besoin métier

Parties prenantes identifiées :

- Direction commerciale
- Direction financière
- Équipe data / IT

- Direction générale

Besoin métier principal :

Disposer d'une prévision fiable des ventes à court terme (J+7), simple à interpréter et utilisable opérationnellement.

Hiérarchisation des besoins :

1. Fiabilité et robustesse de la prévision
2. Simplicité d'interprétation
3. Rapidité de mise en œuvre
4. Facilité d'intégration dans des outils existants

Contraintes :

- Horizon court (J+7)
 - Données agrégées (pas de granularité magasin)
 - Solution légère, peu coûteuse
 - Absence de contraintes réglementaires fortes (données anonymisées)
-

2. Audit de la solution data existante

2.1 Solution actuelle ou proposée

Situation initiale :

- Pas de modèle de prévision automatisé existant.
- Exploitation des données principalement via des analyses historiques.

Outils et technologies proposées :

- Python
- pandas / numpy

- scikit-learn
- FastAPI
- Streamlit

Pipeline de données :

1. Chargement des données historiques
 2. Construction de features temporelles
 3. Création de variables d'historique (lags, moyennes glissantes)
 4. Entraînement du modèle
 5. Exposition via API / application
-

2.2 Évaluation de l'adéquation aux besoins

Critères d'analyse :

- Performance prédictive
- Interprétabilité
- Simplicité de maintenance
- Coût technique

Limites identifiées :

- Pas de prévision automatisée initialement
- Forte dépendance à l'analyse humaine
- Absence de solution réutilisable en production

Conclusion de l'audit :

Une solution de prévision simple et interprétable est pertinente pour répondre au besoin métier.

3. Identification d'une solution technique cible

Comparatif des approches

Approche	Avantages	Inconvénients
Modèle naïf	Très simple	Peu précis
Régression linéaire	Interprétable, robuste	Relations linéaires
Random Forest	Non-linéaire	Complexité, moins interprétable

👉 La régression linéaire a été retenue comme solution cible.

Architecture cible (simplifiée)

- Modèle ML entraîné et sérialisé (joblib)
 - API FastAPI pour exposer la prédiction
 - Application Streamlit pour la démonstration
 - CI pour automatiser les tests
-

Facteurs clés de succès

- Features explicables
 - Pipeline reproductible
 - Validation métier des entrées
 - Monitoring léger des prédictions
-

4. Appui stratégique et méthodologique

4.1 Démarche projet

Méthodologie inspirée de **CRISP-DM** :

1. Compréhension du besoin métier
2. Compréhension des données
3. Préparation des données
4. Modélisation
5. Évaluation
6. Mise en production légère

Roadmap synthétique :

- Semaine 1 : exploration & features
- Semaine 2 : modélisation & évaluation
- Semaine 3 : API, tests, déploiement Streamlit

Estimation des ressources nécessaires

Temps

La réalisation du projet a nécessité environ 16 jours de travail, répartis de la manière suivante :

- Analyse du besoin métier et compréhension des données : 3 jours
- Préparation des données et feature engineering : 5 jours
- Modélisation et évaluation : 4 jours
- Déploiement (API, application Streamlit) et tests : 3 jours
- Documentation, reporting et finalisation : 1 jour

Coûts

Le projet a été conçu dans une logique de coût maîtrisé :

- Outils open source (Python, scikit-learn, FastAPI, Streamlit) : 0 €
- Hébergement Streamlit Community Cloud : 0 €

Le coût total du projet est donc très faible, rendant la solution accessible et facilement reproductible.

Infrastructure

L'infrastructure nécessaire est volontairement légère :

- Poste de travail standard (CPU)
- Environnement Python local
- Dépôt GitHub pour le versioning
- API FastAPI pour l'exposition du modèle
- Application Streamlit pour la démonstration
- Intégration continue via GitHub Actions

4.2 Aide à la prise de décision

Risques :

- Extrapolation hors périmètre temporel
- Rupture de comportement de vente

Opportunités :

- Automatisation de la prévision
- Aide directe à la décision métier

Indicateurs de succès :

- MAE / RMSE
 - Stabilité des prédictions
 - Temps de réponse de l'API
-

5. Contrôle et suivi du projet

5.1 Tableau de bord de pilotage

Indicateurs suivis :

- Performance ML (MAE, RMSE)
- Latence des prédictions
- Nombre de prédictions effectuées

Reporting :

- Tests automatisés à chaque mise à jour
 - Logs de prédictions consultables
-

5.2 Outils et process de suivi

- pytest pour les tests
 - GitHub Actions pour la CI
 - Logs JSON pour le monitoring
 - GitHub pour la collaboration
-

6. Conclusion & recommandations

Décisions prises

- Le choix d'un modèle simple et interprétable a été privilégié afin de maximiser l'adoption métier et de limiter les coûts de maintenance
- Mise en production via API
- Déploiement d'une application de démonstration

Perspectives d'évolution

- Réentraînement périodique
- Ajout de données prix/promotions
- Déclinaison à une granularité plus fine

Prochaines étapes

- Industrialisation du réentraînement
- Suivi de dérive des données
- Intégration dans un outil métier

Bilan critique

- La solution répond efficacement au besoin métier initial dans un cadre simplifié. Néanmoins, elle reste sensible aux ruptures de comportement et nécessite un suivi régulier pour garantir la stabilité des prédictions
-