

# **Cahier des charges**

## **Valentin - Joel - Ivan - Asma**

### **1.introduction :**

- **Contexte et objectifs:**

Nous sommes confrontés à un projet visant à fournir des conseils et des solutions en matière de stratégie d'entreprise, avec une orientation spécifique vers l'utilisation des données. Dans ce cadre, notre objectif principal est d'accompagner notre client, une grande entreprise du secteur de l'e-commerce en ligne, dans l'exploration et l'exploitation de ses données clients pour améliorer ses performances et prendre des décisions plus éclairées. L'essor de l'intelligence artificielle offre des opportunités uniques pour extraire des insights pertinents à partir de ces données, ce qui constitue le cœur de notre mission.

- **Périmètre du Projet**

Le projet se concentre sur les ventes de produits électroniques, incluant l'analyse des données de vente, les tendances d'achat et les comportements des clients.

Inclus :

- Analyse des données de vente.
- Développement de modules d'apprentissage automatique.
- Création de dashboards de datavisualisation.
- Recommandations pour l'utilisateur.

Exclus :

- Analyse des ventes d'autres sections.
- Modules non liés aux produits électroniques.
- Intégration complète au-delà du prototype.
- Support post-projet après démonstration.
- utilisation des données autres que celles fournies par le client.

Cela permet de fournir des analyses détaillées et des solutions spécifiques, démontrant notre expertise et valeur ajoutée.

## **2. Description du Projet**

- **Problématique et enjeux :**

L'entreprise cliente, spécialisée dans l'e-commerce, dispose d'un vaste ensemble de données de ventes de produits électroniques. Cependant, ces données ne sont pas encore exploitées de manière optimale pour en tirer des insights précieux. La problématique principale est de transformer ces données brutes en informations exploitables pour améliorer la prise de décision et la stratégie commerciale. Les enjeux pour l'organisation incluent :

1. Amélioration des performances de vente : Identifier les tendances et prédire les ventes pour mieux gérer les stocks et les campagnes marketing.
2. Compréhension approfondie des clients : Analyser le comportement des clients pour personnaliser les offres et améliorer l'expérience client.
3. Optimisation des ressources : Utiliser les insights pour optimiser les processus internes et allouer les ressources de manière plus efficace.

- **Objectifs détaillés :**

1. Améliorer les prédictions de ventes :
  - Développer des modèles d'apprentissage automatique pour prédire les ventes futures de produits électroniques.
  - Identifier les facteurs influençant les variations des ventes.
2. Analyser le comportement des clients :
  - Segmenter les clients en fonction de leurs comportements d'achat.
  - Identifier les clients les plus fidèles et comprendre leurs préférences.
3. Optimiser les campagnes marketing :
  - Utiliser les insights pour créer des campagnes marketing ciblées et efficaces.
  - Mesurer l'impact des campagnes sur les ventes et ajuster les stratégies en conséquence.
4. Créer des dashboards de datavisualisation :
  - Développer des dashboards interactifs pour visualiser les données de vente et les prédictions de manière intuitive.
  - Fournir des outils de visualisation permettant aux décideurs de suivre les performances et d'identifier rapidement les opportunités et les problèmes.

### 3. Données

- **Sources de données** : Notre principale source de données est un dataset qui prend la forme d'un fichier .csv, récupéré depuis un lien WeTransfer.
- **Types de données** : Les données ne sont pas structurées dans leur intégralité. Il y a des données manquantes, des doublons, et des données de type date, entre autres, qui nécessitent un prétraitement.
- **Qualité des données** : Aucune erreur n'a été générée par le dataset sur Power BI, mais les colonnes category\_code et brand ont plus de 20% de valeur manquantes( avec 8% de valeurs manquantes dans les deux colonnes au même temps) , La colonne user\_session aussi a elle-aussi quelques lignes sans valeur. Bien que ça ne semble pas illogique pour un jeu de données sur une vente de produits électroniques, on observe néanmoins une grande disparité entre la moyenne et la valeur médiane, ce qui suggère qu'il y aurait ce qui serait normalement considéré comme des valeurs aberrantes.

### 4. Fonctionnalités et Exigences

- **Exigences fonctionnelles** :
  1. Tableaux de bord interactifs :
    - Visualisation des données de ventes par catégorie de produits, période de temps et marque.
    - Graphiques dynamiques permettant de suivre les tendances de vente et les prédictions.
    - Filtrage et exploration des données pour une analyse approfondie.
  2. Rapports analytiques :
    - Rapports automatisés sur les performances de vente, les prévisions, et les insights clés.
    - Analyse comparative des ventes par rapport aux objectifs fixés.
  3. Modèles de prédiction :
    - Développement de modèles d'apprentissage automatique pour prédire les ventes futures.

- **Exigences non-fonctionnelles :**

1. Performance :

- Temps de réponse rapide pour les requêtes et la visualisation des données.
- Capacité à traiter de grandes quantités de données sans compromettre les performances.

2. Sécurité :

- Protection des données clients et sensibles conformément au RGPD (En anonymisant l'identifiant des utilisateurs)
- Contrôle d'accès pour limiter l'accès aux informations confidentielles.

3. Scalabilité :

- Capacité à gérer une augmentation du volume de données sans perte de performance.

- **Critères de réussite :**

1. Précision des prévisions :

- Comparaison des prévisions avec les ventes réelles pour évaluer la précision des modèles.
- Atteinte d'un niveau de précision défini dans les prévisions.

2. Utilisation et Adoption :

- Taux d'adoption par les utilisateurs finaux.
- Feedback positif sur la convivialité et l'utilité des outils développés.

3. Amélioration des Performances :

- Augmentation des ventes de produits électroniques après la mise en œuvre des recommandations basées sur l'analyse des données.
- Réduction du temps passé sur les tâches analytiques grâce à l'automatisation des rapports et des prévisions.

## **5. Technologie et Architecture**

- **Environnement technique :** Pour ce qui est preprocessing, nettoyage de la donnée, et réseaux de neurones, on utilise Python sous plusieurs supports, comme Visual Studio, Google Collab, ou en local. On aura recours à plusieurs bibliothèques, On utilise également Power BI pour le traitement et surtout la visualisation.

- **Architecture** : Découpage et traitement des données et de leur qualité sur Python, visualisation avec matplotlib et Power Bi. Modèle d'apprentissage sur Python.

## 6. Sécurité et Confidentialité

- **Mesures de Sécurité**
  - Politiques de Sécurité : Mise en place de politiques strictes pour protéger les données.
  - Cryptage des Données : Utilisation de cryptage pour sécuriser les données en transit et au repos.
  - Accès Restreint : Contrôle d'accès basé sur les rôles pour limiter l'accès aux informations sensibles aux seuls utilisateurs autorisés.
- **Conformité Réglementaire**
  - Conformité aux Régulations : Respect des réglementations et normes applicables, telles que le RGPD pour les données personnelles en Europe.
  - RGPD : Anonymisation des données, obtention du consentement explicite des utilisateurs et mise en place de mécanismes pour l'exercice des droits des utilisateurs (accès, rectification, suppression des données).

## 7. Organisation et Gestion du Projet

- **Équipe Projet : Valentin - Joel - Lazare - Asma**
- **rôles et responsabilités:**

**Valentin | Lazare:** Data Analyst

**Joel:** Project Manager, Data scientist

**Asma:** Data Productrice

- **Planning** : Calendrier détaillé avec les phases et les jalons du projet.

5 jours:

- Jour 1: Extraction, Lecture, Identification des valeurs
- Jour 1: Preprocessing
- Jour 2: Implémentation du modèle
- Jour 2: Test du résultat

- Jour 3: Création de l'interface l'application
- Jour 3: Implémentation de l'API pour client-serveur
- Jour 4: Test Application Complète
- Jour 4: Déploiement

Après

- Maintenance ou Mise à Jour du Modèle
- **Budget** : estimation des coûts du projet et allocation des ressources.

Supposons que les salaires journaliers des membres de l'équipe soient les suivants :

- Valentin : 200 €/jour
- Lazare : 200 €/jour
- Joël : 300 €/jour
- Asma : 150 €/jour

Alors, les coûts totaux du projet seraient :

- Valentin (4 jours) : 4 jours \* 200 €/jour = 800 €
- Lazare (4 jours) : 4 jours \* 200 €/jour = 800 €
- Joel (4 jours) : 4 jours \* 300 €/jour = 1200 €
- Asma (4 jours) : 4 jours \* 150 €/jour = 600 €

Le coût total du projet serait la somme de ces coûts individuels :

800 € (Valentin) + 800 € (Lazare) + 1200 € (Joël) + 600 € (Asma) = 3400 €

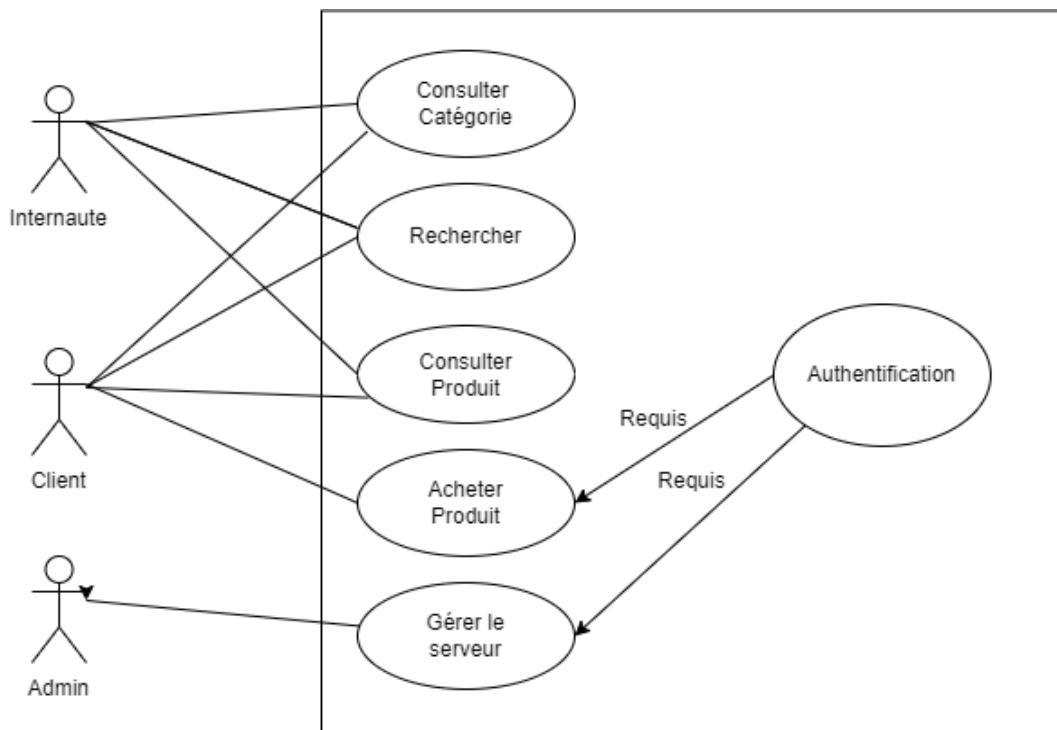
## 8. Risques et Contingences

- **Identification des risques** :
1. Risques Techniques :
    - Défaillance des modèles Prédicatifs : Les modèles d'apprentissage automatique peuvent ne pas fournir des prédictions précises.
    - Problèmes d'Intégration : Difficultés à intégrer les modules dans l'infrastructure existante du client.
  2. Risques Opérationnels :
    - Résistance au Changement : Les utilisateurs peuvent être réticents à adopter de nouveaux outils et méthodes.

- Formation Insuffisante : Les utilisateurs finaux peuvent ne pas recevoir une formation adéquate pour utiliser la nouvelle plateforme efficacement.
3. Risques de Sécurité :
- Violations de Données : Risque de piratage ou de fuite de données sensibles.
  - Non-conformité RGPD : Risques de non-respect des réglementations sur la protection des données, entraînant des sanctions légales.

- **Plan de mitigation :**

1. Techniques : Tests rigoureux des modèles, intégration progressive, procédures de débogage.
2. Opérationnels : Plan de gestion du changement, sessions de formation complètes.
3. Sécurité : Mesures de sécurité renforcées, conformité RGPD, surveillance et audits réguliers.



**Diagramme de cas d'utilisation globale**

