

ANALYSER UN BESOIN/ UNE DEMANDE EN IA

MSPR 1

Project by

Shiyun TANG
Belkis COSKUN
Mohamed ELHAFA
Abdelkim BENKIRANE

School

EPSI B3 DEVIA

Our Topic

**How Does the Use of
Artificial Intelligence
Improve the Efficiency of
Radiological Medicine?**

Tool

**AI Platform for Medical
Image Analysis**

Sommaire

Introduction	1
---------------------	----------

Cahier de charges	2
--------------------------	----------

• Présentation d'Amazing	2
--------------------------	---

• Problématique	2
-----------------	---

• Analyse des données disponibles (Audit initial)	3
---	---

• Objectifs fonctionnels	3
--------------------------	---

• Stratégie Marketing pilotée par IA	3
--------------------------------------	---

• KPI de Suivi du Projet	3
--------------------------	---

• Roadmap et Étapes Clés du Projet	3
------------------------------------	---

• Architecture Data & Processus Métier	4
--	---

• Besoin (Cadrage non-technique)	5
----------------------------------	---

• Audit des Données Disponibles	6
---------------------------------	---

• Étude de Faisabilité : Risques et Contraintes	7
---	---

• Justification du Choix de Projet Data Science	
---	--

• Évaluation RGPD & Actions Recommandées	
--	--

Audit des données disponibles	8
Analyse de faisabilité et contraintes	9
Justification de la solution IA retenue	11
Veille technologique, sectorielle, concurrentielle et réglementaire	
Indicateurs de performance (KPI) et plan de suivi	
Étude des outils et technologies retenus	
Roadmap du projet	
Conformité RGPD et gouvernance des données	
Évaluation des moyens prévisionnels et estimation budgétaire	
Note de faisabilité synthétique	
Conclusion	

Introduction

La révolution numérique redéfinit en profondeur les pratiques commerciales, et l'intelligence artificielle (IA) s'impose comme un levier stratégique majeur dans cette transformation. Dans un contexte où les attentes des consommateurs évoluent rapidement, l'IA offre des perspectives inédites pour mieux comprendre les comportements d'achat, personnaliser les offres, et optimiser l'efficacité des actions marketing.

La problématique centrale que nous abordons est donc la suivante : comment exploiter l'intelligence artificielle pour personnaliser l'expérience client et stimuler les ventes, tout en garantissant l'interprétabilité, la conformité réglementaire et l'intégration opérationnelle de la solution ?

Ce cahier des charges s'inscrit dans cette dynamique d'innovation, en se concentrant sur l'usage de l'intelligence artificielle pour améliorer les performances commerciales de la marque Amazing Basics, en particulier sur les catégories en perte de vitesse telles que le divertissement. L'objectif principal est de concevoir une solution data-driven, fondée sur des modèles prédictifs et de recommandation, afin d'augmenter la pertinence des campagnes marketing, renforcer la fidélisation client, et maximiser le retour sur investissement.

L'IA représente une opportunité stratégique pour Amazing : en combinant puissance analytique et adaptation en temps réel, elle permet de dépasser les approches traditionnelles basées sur des segments fixes ou des promotions génériques. Cependant, cette transformation soulève également des enjeux complexes : gouvernance des données, respect du RGPD, lisibilité des modèles, et appropriation par les équipes non techniques.

Ce cahier des charges propose une démarche structurée pour cadrer, concevoir et déployer une solution IA sur mesure. Il est organisé autour de plusieurs dimensions clés : analyse des besoins métiers, audit des données disponibles, choix méthodologiques, planification, contraintes, faisabilité et conformité. L'ambition est d'apporter aux décideurs une vision claire, pragmatique et orientée résultats, pour tirer pleinement parti du potentiel de l'intelligence artificielle dans un cadre maîtrisé et responsable.



Présentation de l'Entreprise : Amazing

Amazing est une marketplace en ligne internationale qui propose une très large gamme de produits allant de l'électronique à la mode, en passant par les accessoires pour la maison, les produits de beauté, et bien d'autres catégories. L'entreprise est un acteur mondial majeur du e-commerce, présent dans de nombreux pays, avec une plateforme performante, une logistique optimisée, et un modèle axé sur la satisfaction client. Une partie significative de son chiffre d'affaires provient de sa marque propre : Amazing Basics, qui regroupe des produits simples, pratiques et à bas prix.

Cette marque a pour objectif de proposer des articles essentiels à tous les types de consommateurs. Cependant, ces derniers mois, Amazing a constaté une diminution de l'engagement client et une baisse des ventes, notamment dans la catégorie "divertissement". Cette situation s'explique par plusieurs facteurs : l'inflation, une concurrence agressive, une saturation du marché, ainsi qu'une faible personnalisation de l'offre actuelle.

Problématique

Malgré un positionnement solide et une base de clients massive, Amazing fait face à un recul de 15 % des ventes sur les produits Amazing Basics au cours des deux derniers trimestres. Cette baisse est accentuée sur les produits liés au divertissement et aux accessoires technologiques, autrefois piliers de la croissance.

Les initiatives marketing classiques (newsletters génériques, promotions saisonnières, etc.) ne suffisent plus à stimuler les ventes. Il est aujourd'hui nécessaire d'évoluer vers des actions plus ciblées, personnalisées et intelligentes. L'objectif est de tirer parti des données clients et des technologies d'IA pour restaurer la croissance et améliorer significativement les ventes.

L'enjeu principal est donc : comment exploiter les données internes de manière intelligente afin de relancer les ventes de la marque Amazing Basics, en optimisant le ciblage marketing, l'expérience utilisateur et la prédiction des besoins client ?



Analyse des Données Disponibles

L'équipe Data d'Amazing met à disposition une grande quantité de données, structurées dans un Data Lake cloud. Ces données couvrent toutes les dimensions de l'expérience client :

- Catalogue des produits : noms, prix, catégories, stock, descriptions.
- Commandes clients : produits achetés, date, fréquence, valeurs paniers.
- Comportement de navigation : pages visées, temps de session, clics, rebonds.
- Base clients anonymisée : données d'inscription, historique, segments.
- Actions marketing passées : promotions, newsletters, codes réductions.

Ces données sont riches et multi-sources. Elles permettent d'imaginer des modèles complexes comme :

- Des systèmes de recommandation personnalisée (filtrage collaboratif, deep learning)
- Des modèles de scoring clients (segmentation RFM, clustering)
- Des prédictions de ventes selon les périodes, les comportements ou les profils

Objectifs Détaillés du Projet

Augmentation du chiffre d'affaires de la marque Amazing Basics de 25 % d'ici 12 mois, via des recommandations intelligentes et des actions ciblées

- Amélioration du taux de conversion des visiteurs en acheteurs (+10 %)
- Hausse du panier moyen (+8 %) grâce à la vente croisée à travers des suggestions IA
- Optimisation du retour sur investissement marketing (ROI campagnes +15 %)
- Réduction du taux de rebond de 20 % sur les pages produits

Pour cela, la solution doit :

- **Identifier les profils d'utilisateurs types et leurs besoins**
- **Détecter les produits sous-performants et ceux à fort potentiel**
- **Personnaliser les recommandations produits et les e-mails**
- **Prévoir la demande pour adapter les prix et les stocks**

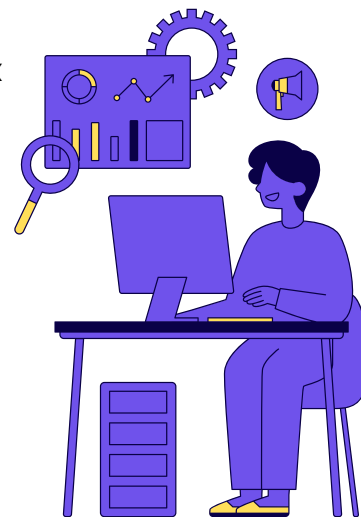


Stratégie Marketing Pilotée par l'IA

La relance des ventes passera par une réorientation du marketing vers la personnalisation :

- Segmentation IA : découper les clients en groupes comportementaux (jeunes technophiles, parents actifs, clients inactifs, etc.)
- Campagnes ciblées : envoi de newsletters et offres personnalisées à fort taux de réponse
- Recommandations temps réel : sur le site ou par email, selon l'historique utilisateur
- Dynamic pricing : moduler les prix selon le moment ou le profil client
- Cross-selling & up-selling automatisé : suggestion de produits associés

Cette stratégie sera testée avec des groupes de clients et optimisée selon les résultats (A/B testing).



—● KPI de Suivi du Projet

- ↑ Taux de conversion (avant/après IA)
- ↑ Panier moyen
- ↑ Taux de clics sur recommandations personnalisées
- ↑ Taux d'ouverture des campagnes emails ciblées
- ↑ Évolution des ventes par catégorie Amazing Basics



Roadmap et Étapes Clés du Projet ●—

La feuille de route du projet a été pensée de manière progressive et itérative, pour garantir à la fois rigueur méthodologique et flexibilité d'adaptation. Elle se structure autour de cinq grandes phases allant de l'analyse initiale jusqu'au déploiement opérationnel et à l'amélioration continue. Chaque étape est définie avec des objectifs clairs, des livrables concrets, et des délais réalistes tenant compte des ressources disponibles, tant du côté de l'équipe data que de celle du commanditaire. Cette roadmap permet d'assurer un suivi fluide du projet, de prioriser les actions critiques, et d'assurer une montée en valeur incrémentale au fil du temps..

Architecture Data & Processus Métier

Étapes clés à représenter dans un schéma métier :

- Collecte des données clients, navigation et produits
- Traitement, nettoyage et agrégation dans le Data Lake
- Modélisation (modèles IA pour recommandation, scoring, prédiction)
- Déclenchement des actions marketing (emails, recommandations web)
- Suivi des KPIs en temps réel via dashboard BI



—● Conformité RGPD et Éthique

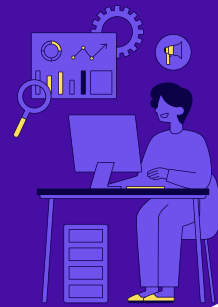
- Toutes les données utilisées doivent être anonymisées et stockées selon les normes européennes
- Le client doit pouvoir accéder, rectifier ou supprimer ses données sur demande
- L'algorithme ne doit pas prendre de décisions discriminatoires (ex. par sexe ou origine)
- Ajout de mécanismes de contrôle humains (interprétabilité des modèles)
- Formation de l'équipe marketing sur les implications légales



—● Outils et Technologies Recommandées

- Data Engineering : AWS S3, AWS Glue, Snowflake, Airflow
- Modélisation IA : Python, Scikit-learn, XGBoost, TensorFlow, LightFM
- Visualisation : Power BI, Tableau, Streamlit pour dashboard interactif
- Déploiement : Docker, FastAPI, CI/CD GitHub Actions
- Surveillance : Prometheus, Grafana pour monitoring

Besoin et objectifs opérationnels



Ce projet vise à répondre à un besoin clairement exprimé par les équipes d'Amazing : relancer les ventes de la marque Amazing Basics grâce à une meilleure connaissance client et à une personnalisation accrue de l'expérience d'achat. L'objectif n'est pas uniquement technique, mais avant tout opérationnel : proposer une solution intelligente, simple à exploiter par les équipes marketing, et intégrée aux outils existants. Il s'agit d'allier innovation en intelligence artificielle et accessibilité métier.



—● Audit des données disponibles

Avant toute phase de modélisation, une évaluation rigoureuse de la qualité, de la pertinence et de la complétude des données disponibles est nécessaire. Cet audit permet d'identifier les forces et les limites des sources actuelles (navigation, ventes, campagnes marketing), et de définir les actions préalables indispensables pour garantir la fiabilité des analyses et des modèles à venir.

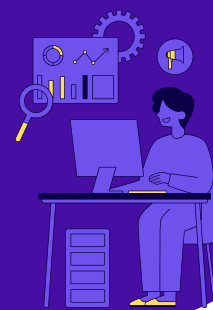
—● Analyse de faisabilité, contraintes et risques

La mise en œuvre du projet doit tenir compte de plusieurs contraintes techniques, humaines et réglementaires. Une analyse de faisabilité permet d'anticiper les principaux risques liés au calendrier, à la complexité des modèles ou encore à la conformité RGPD. Cette évaluation conditionne la réussite du projet, en assurant un équilibre entre ambition technologique et réalisme opérationnel.

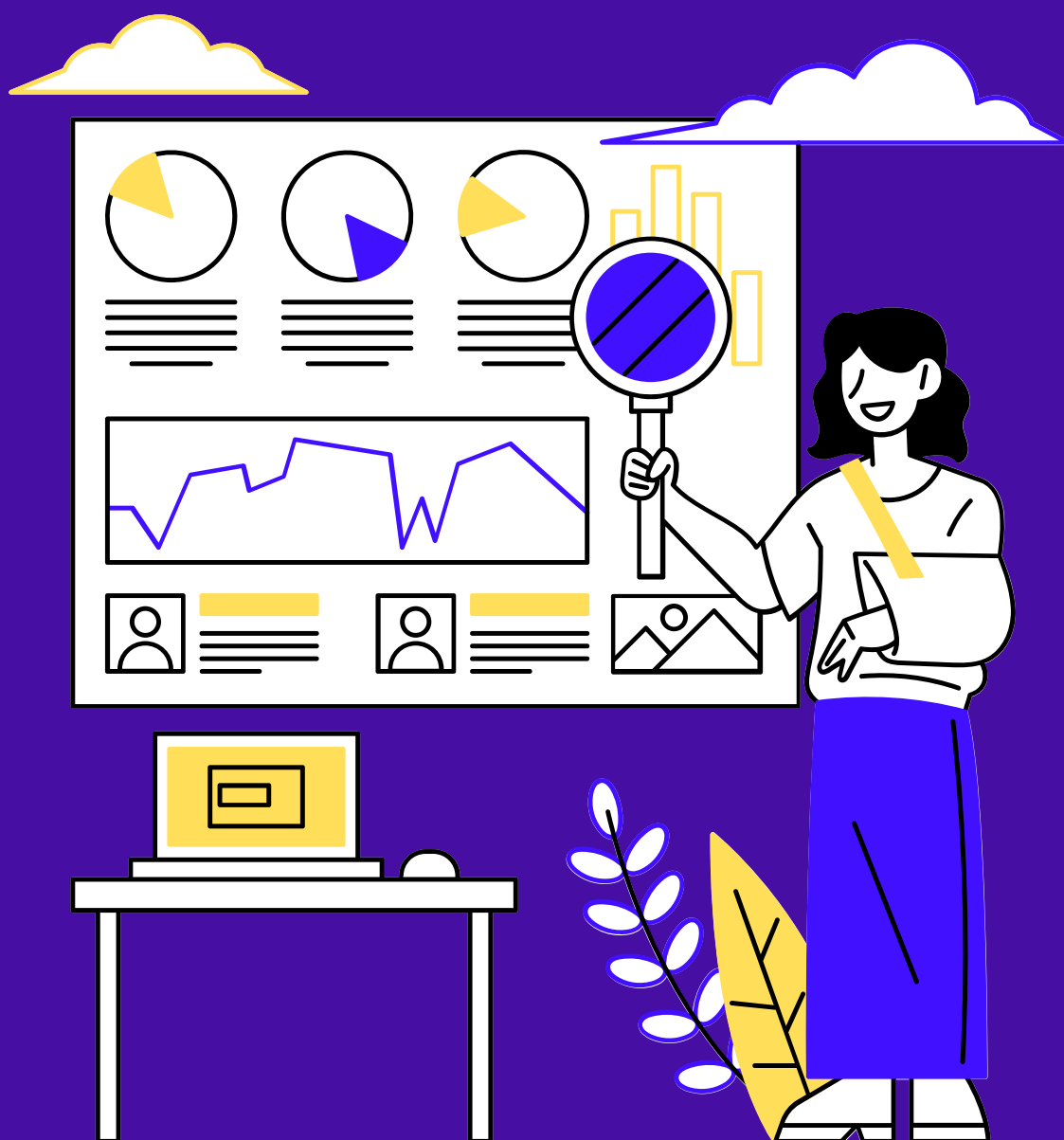
—● Pertinence de la solution proposée

Le recours à une approche data science se justifie pleinement au regard des enjeux identifiés. Cette démarche permet de répondre de manière ciblée et personnalisée aux attentes des clients tout en optimisant les actions marketing. Le projet proposé s'inscrit dans une logique de création de valeur mesurable, durable et adaptée aux nouvelles dynamiques du marché.

Conformité RGPD et gouvernance des données



L'ensemble du projet sera conduit dans le strict respect du Règlement Général sur la Protection des Données (RGPD). Au-delà des obligations légales, la gouvernance des données constitue un levier de confiance auprès des utilisateurs. Des actions concrètes seront mises en place pour garantir la transparence, la sécurité et l'équité des traitements de données tout au long du projet.



Audit des données disponibles

Le jeu de données utilisé pour ce projet a été extrait de la plateforme Rees46. Il regroupe plusieurs mois d'activités e-commerce simulées (d'octobre 2019 à avril 2020), représentant les interactions entre utilisateurs et produits sur un site de vente en ligne.

Ces données, au format .csv.gz, représentent un volume total de plus de 15 Go et sont particulièrement riches pour analyser les comportements utilisateurs à grande échelle.

Schéma des données

Colonne	Description
`event_time`	Date et heure de l'événement (UTC)
`event_type`	Type d'événement : `view`, `cart`, `purchase`, etc.
`product_id`	Identifiant unique du produit
`category_id`	Identifiant numérique de la catégorie produit
`category_code`	Nom textuel de la catégorie (ex : electronics.smartphones)
`brand`	Marque du produit (peut être manquante ou vide)
`price`	Prix du produit en float
`user_id`	Identifiant unique de l'utilisateur (pseudonymisé)
`user_session`	Identifiant temporaire de session utilisateur

• QUALITÉ ET LIMITES DES DONNÉES

- Richesse comportementale : les événements permettent de reconstituer des parcours utilisateurs : consultation, ajout au panier, achat.
 - Sessions claires : le champ user_session permet d'isoler les sessions d'achat, utile pour la recommandation basée sur le contexte.
 - Pas de profil client détaillé (âge, sexe, lieu, etc.) → la segmentation sera donc comportementale uniquement.
- Données de prix bien renseignées, mais certaines valeurs peuvent être nulles ou aberrantes (prix = 0 ou très élevés).
 - Catégories parfois absentes (category_code manquant), ce qui complique certaines analyses sur les familles de produits.
 - Pas de données marketing disponibles (newsletters, campagnes email, codes promo).

• DONNÉES UTILES SELON LES OBJECTIFS IA

OBJECTIF IA	DONNÉES UTILISABLES
Recommandation produits	user_id, product_id, event_type, user_session, price
Segmentation comportementale	Fréquence des sessions par user_id, taux de conversion, récurrence
Prédiction des ventes	event_time, event_type=purchase, price, category_code

ce jeu de données ne contient ni informations personnelles ni données marketing, il est parfaitement adapté à la mise en place :

- d'un système de recommandation personnalisé (type collaborative filtering),
- d'une segmentation comportementale (clustering RFM ou par sessions),
- d'une prévision des ventes par produit ou catégorie, en exploitant les données temporelles (event_time), les prix et les achats réalisés.

Il conviendra de nettoyer les données, filtrer les sessions valides, gérer les colonnes manquantes et agréger les événements par utilisateur ou session pour entraîner des modèles efficaces.



Analyse de faisabilité et contraintes

La faisabilité du projet repose sur une évaluation rigoureuse des contraintes techniques, humaines et réglementaires. Cette analyse permet d'anticiper les risques potentiels et d'aligner la solution IA avec les capacités réelles de l'organisation et son écosystème numérique.

• CONTRAINTES TECHNIQUES

Infrastructure nécessaire : le traitement de données volumineuses (>15 Go) et l'entraînement de modèles IA nécessitent une infrastructure scalable. Amazing dispose déjà d'un Data Lake cloud, ce qui est un atout.

Prétraitement important : les données brutes (fichiers .csv.gz) exigent des étapes de nettoyage, d'agrégation, et de gestion des valeurs manquantes avant toute modélisation.

Absence de données contextuelles (stock, marketing, données client enrichies) : la prédiction reste limitée aux comportements utilisateurs enregistrés.

Outils nécessaires : Python, scikit-learn, LightFM, XGBoost, ou encore FastAPI pour le déploiement.

• CONTRAINTES HUMAINES

Adoption par les équipes métiers : les équipes marketing devront être formées à l'exploitation des résultats IA (tableaux de bord, modèles de recommandation, scores de segmentation).

Compréhension des modèles : certains algorithmes (Random Forest, XGBoost) peuvent être vus comme des "boîtes noires". Il sera nécessaire de fournir des visualisations ou explications simplifiées pour faciliter leur appropriation.

Résistance au changement : il faut anticiper une phase d'accompagnement pour que les équipes passent d'un marketing générique à un marketing basé sur la donnée.



• **CONTRAINTES RÉGLEMENTAIRES (RGPD)**



Les données utilisées sont pseudonymisées (user_id, user_session) et ne contiennent pas d'informations personnelles directes, ce qui limite les risques.

Cependant, l'utilisation de ces données à des fins de profilage nécessite :

- Une transparence sur les traitements
- Une évaluation des biais possibles
- L'intégration de mécanismes de contrôle (droits d'accès, rectification, suppression)

L'analyse de la conformité devra être validée avec un DPO (Data Protection Officer) en cas de mise en production réelle.

• **ÉVALUATION DES RISQUES (MÉTHODE EY)**

OBJECTIF IA	RISQUE POTENTIEL	RECOMMANDATION
Recommandation produits	Discrimination indirecte (prix, catégorie)	Vérifier les biais, tester les segments
Robustesse	Modèles sensibles à la saisonnalité ou au bruit	Intégrer du lissage, ou des modèles temporels
Explicabilité	Modèles complexes (XGBoost, Forest) peu lisibles	Utiliser SHAP, visualisation des scores
Confiance métier	Perception de la "boîte noire", rejet possible	Sensibilisation, restitution pédagogique

Le projet est techniquement réalisable, à condition d'encadrer rigoureusement le nettoyage des données, de choisir des modèles interprétables et de mettre en place un accompagnement humain. L'ensemble du dispositif devra respecter les exigences du RGPD, avec une attention particulière portée à l'éthique algorithmique et à la transparence des traitements.

Justification de la solution IA retenue

Afin de répondre efficacement aux besoins exprimés par l'entreprise Amazing, nous avons identifié trois axes fonctionnels prioritaires : la personnalisation de l'expérience utilisateur, la segmentation comportementale de la clientèle, et la prévision de la demande. Pour chacun de ces axes, nous avons sélectionné des modèles d'intelligence artificielle adaptés, en tenant compte des contraintes techniques du projet, de la structure des données disponibles, de la robustesse des algorithmes et de leur degré d'interprétabilité. Cette section présente les modèles choisis, les raisons de leur sélection, ainsi qu'une justification des alternatives écartées.

• RECOMMANDATION PERSONNALISÉE

La baisse d'engagement constatée sur la marque Amazing Basics appelle une stratégie de personnalisation renforcée. Un système de recommandation permet de proposer aux utilisateurs des produits correspondant à leurs préférences et comportements passés, favorisant ainsi la conversion et la fidélisation.

Modèle retenu : LightFM

Nous avons choisi l'algorithme LightFM, un système de recommandation hybride combinant :

- le filtrage collaboratif (basé sur les interactions entre utilisateurs et produits),
- et le filtrage par contenu (caractéristiques produits : catégorie, prix, marque).

LightFM est particulièrement adapté à notre dataset, qui contient des événements de type view, cart, purchase, associés à des identifiants produits, utilisateurs, catégories et prix.

Avantages principaux :

- Gère efficacement le problème du cold-start (nouveaux produits ou utilisateurs peu actifs).
- Permet d'exploiter les attributs produits en plus des historiques d'interactions.
- Offre des performances satisfaisantes avec un temps d'entraînement raisonnable.

Pourquoi pas d'autres modèles ?

- Les approches par réseaux de neurones (type Autoencoders ou GRU4Rec) sont plus complexes, nécessitent davantage de données nettoyées et de puissance de calcul.
- Le filtrage collaboratif pur (ex. SVD) n'utilise pas les attributs produits, limitant la précision dans notre contexte où certaines catégories sont mal renseignées.



• SEGMENTATION COMPORTEMENTALE

La compréhension des profils clients est essentielle pour adapter les actions marketing. En l'absence de données socio-démographiques, une segmentation basée sur le comportement d'achat s'impose.

Modèle retenu : K-Means sur une base RFM

Nous avons retenu une approche classique de segmentation RFM (Récence, Fréquence, Montant), sur laquelle nous appliquons le clustering K-Means. Cette méthode permet de regrouper les clients selon leur activité d'achat et de détecter :

- des profils fidèles et actifs,
- des clients à réactiver,
- ou encore des profils occasionnels.

Avantages principaux :

- Algorithme rapide, facilement interprétable et applicable à grande échelle.
- Représentation visuelle claire des clusters.
- Actionnable directement par les équipes marketing.

Pourquoi pas d'autres modèles ?

- DBSCAN est plus adapté à la détection d'anomalies qu'à la segmentation marketing classique.
- Les modèles de mélange gaussien (GMM) sont plus lents et sensibles au bruit, et peu lisibles pour un usage métier.
- Le clustering hiérarchique est difficile à appliquer sur des données volumineuses comme les nôtres.

• PRÉVISION DES VENTES

Anticiper la demande est un enjeu stratégique pour optimiser les stocks, planifier les campagnes et ajuster les prix. La nature temporelle de nos données justifie une approche de prévision multivariée des ventes.

Modèles retenus :

- **XGBoost (modèle principal)**
- **Random Forest (modèle de benchmark)**

a) XGBoost

XGBoost est un algorithme de boosting d'arbres de décision, reconnu pour son efficacité sur des données tabulaires complexes.

Avantages principaux :

- Gère bien les relations non linéaires entre les variables (ex. : influence du prix selon la catégorie).
- Offre une grande robustesse même avec des données bruitées ou incomplètes.
- Permet d'identifier les variables les plus influentes dans les prédictions (interprétabilité métier).

Pourquoi pas la régression linéaire ?

La relation entre les variables explicatives et le volume des ventes n'est pas linéaire, ce qui limite la pertinence de modèles trop simples.

c) Random Forest

Enfin, nous avons ajouté le modèle Random Forest comme alternative complémentaire à XGBoost. Il constitue un benchmark robuste pour la régression, avec des résultats souvent comparables à XGBoost en termes de précision, mais plus simple à mettre en œuvre.

Avantages principaux :

- Très bon compromis performance / interprétabilité.
- Facilité d'implémentation sans réglage complexe.
- Moins sensible aux hyperparamètres que les modèles de boosting.

Pourquoi l'inclure malgré XGBoost ?

L'objectif est de comparer plusieurs modèles sur les mêmes indicateurs de performance (RMSE, MAE, etc.) afin de valider le modèle le plus adapté en phase de test.

Données utiles selon les objectifs IA

OBJECTIF IA	TYPE DE MODÈLE IA	MODÈLES LES PLUS COURANTS	BRÈVE EXPLICATION
Recommandation produits	Filtrage collaboratif ou contenu	-Matrix Factorization (ALS, SVD) -LightFM -Nearest Neighbors (KNN) -Réseaux de neurones (Autoencoders, Deep Learning)	On prédit les produits qu'un utilisateur pourrait aimer en se basant sur ses interactions passées (clics, achats) ou sur les similarités avec d'autres utilisateurs.
Segmentation comportementale	Clustering non supervisé	-K-Means -DBSCAN -Hierarchical Clustering	On regroupe les clients selon leur comportement (fréquence d'achat, récurrence, panier moyen...) pour adapter les actions marketing.
Prédiction des ventes	Régression supervisée / séries temporelles	-XGBoost / LightGBM / CatBoost -ARIMA / SARIMA -Facebook Prophet - RNN / LSTM (Deep Learning)	On prédit le volume ou le montant des ventes futures selon les dates, catégories, prix, etc.

— La combinaison des modèles retenus — LightFM, K-Means (RFM), XGBoost, Prophet, Random Forest — permet de répondre de manière complète et cohérente aux objectifs du projet. Elle concilie pertinence métier, performance technique, robustesse face aux données réelles, et capacité à être comprise et exploitée par les équipes opérationnelles d'Amazing.

Ce choix réfléchi garantit une solution IA actionnable, évolutive, et intégrée dans une démarche responsable, en cohérence avec les principes d'éthique, d'explicabilité et de conformité posés dès le début du projet.

Veille technologique, sectorielle, concurrentielle et réglementaire

La mise en œuvre d'un projet d'intelligence artificielle ne peut se faire sans une compréhension fine de son environnement externe. Dans un secteur aussi évolutif que le e-commerce, il est essentiel de s'inscrire dans une dynamique d'innovation continue, en intégrant les évolutions technologiques, les usages du marché, les pratiques des concurrents et les contraintes juridiques. Cette veille multidimensionnelle alimente les décisions stratégiques du projet et garantit son alignement avec les standards du secteur.

• VEILLE TECHNOLOGIQUE

Ces dernières années, plusieurs tendances technologiques transforment les usages de l'intelligence artificielle dans les entreprises :

L'IA générative : représentée par des modèles comme GPT-4 ou DALL-E, elle permet de générer automatiquement du contenu (textes, images, recommandations, réponses client, etc.). Dans le e-commerce, elle est utilisée pour créer des fiches produits dynamiques, générer des réponses automatisées aux clients ou produire des recommandations ultra-personnalisées.

Les modèles explicables (XAI – Explainable AI) : avec des outils comme SHAP, LIME ou ELI5, les entreprises peuvent visualiser les contributions des variables dans les décisions des modèles. C'est un enjeu central dans un contexte RGPD où la transparence est primordiale.

L'IA embarquée : cette tendance consiste à intégrer directement les algorithmes d'IA dans des systèmes embarqués (capteurs, objets connectés, terminaux mobiles), afin de proposer des recommandations locales sans dépendance au cloud. Dans le e-commerce, cela permet des expériences client plus fluides et réactives (recommandation instantanée en application mobile, chatbot intégré, etc.).

Le Edge Computing et le MLOps : les workflows MLOps (Machine Learning Operations) deviennent incontournables pour industrialiser les modèles IA. Ils permettent d'automatiser les phases de versioning, tests, déploiement et surveillance des modèles.

• Positionnement du projet :

Notre solution IA s'intègre dans cette dynamique technologique, notamment via :

- l'utilisation d'un système de recommandation hybride interprétable (LightFM),
- la sélection de modèles explicables (XGBoost avec importance des variables, Random Forest avec SHAP),
- l'ouverture possible à des outils de monitoring (Streamlit, Prometheus, etc.) pour accompagner le déploiement.

• VEILLE CONCURRENTIELLE

Pratiques des leaders du secteur: Les grandes plateformes e-commerce exploitent l'IA de manière avancée pour personnaliser l'expérience utilisateur et optimiser la rentabilité.

Amazon : pionnier du filtrage collaboratif à grande échelle, Amazon utilise un système de recommandation combiné à du deep learning. L'algorithme est optimisé en continu à partir de milliards d'interactions. Amazon utilise également le dynamic pricing (variation de prix en temps réel), basé sur la demande, le profil utilisateur et la concurrence.

Netflix : bien qu'opérant dans le secteur du streaming, Netflix est reconnu pour sa recommandation basée sur l'historique utilisateur et ses algorithmes de clustering comportemental, qu'il ajuste pour chaque profil.

Zalando : l'entreprise allemande mise sur la personnalisation du catalogue, en croisant comportement utilisateur et tendances de mode. Elle applique aussi le marketing prédictif pour anticiper les produits à promouvoir auprès de chaque segment.

Spotify et YouTube : bien qu'orientés média, ces géants exploitent des recommandations contextuelles dynamiques, inspirantes pour la structuration d'un moteur de recommandation e-commerce.

• Apport au projet :

Ces benchmarks nous confirment que notre choix d'un moteur de recommandation hybride (LightFM) est cohérent avec les pratiques de référence, tout en restant plus léger à déployer.

• VEILLE SECTORIELLE

Tendances IA dans le e-commerce: Le secteur du commerce en ligne évolue rapidement vers une exploitation intensive des données. Voici les principales tendances qui structurent les stratégies IA des entreprises du secteur :

Personnalisation avancée : elle devient un standard, allant au-delà de la simple recommandation de produit. Il s'agit d'adapter l'ensemble du parcours client (page d'accueil, suggestions, contenu, prix) en fonction du comportement.

Marketing prédictif : les marques utilisent des algorithmes de prédiction pour anticiper les abandons de panier, les comportements de désabonnement, ou encore les besoins latents des utilisateurs. Cela permet de cibler plus intelligemment les campagnes.

Dynamic Pricing : les prix ne sont plus fixes, mais évolutifs, en fonction de la demande, du profil utilisateur, de la saisonnalité ou de l'état du stock. Cette pratique devient de plus en plus courante, notamment dans la mode, la tech ou les voyages.

Segmentation comportementale : grâce à l'IA, les entreprises délaissent les segments sociodémographiques classiques au profit de clusters dynamiques basés sur les interactions, la récurrence ou la fréquence d'achat.

- **Lien avec notre projet :**

Notre solution IA s'aligne parfaitement avec ces tendances sectorielles : segmentation par K-Means, marketing personnalisé via des clusters RFM, et recommandations ciblées en fonction du comportement utilisateur.

- **VEILLE RÉGLEMENTAIRE**

Cadre légal et bonnes pratiques: Le déploiement d'un projet IA, même dans un cadre pédagogique, doit être pensé dans le respect des règles de protection des données personnelles et des principes d'éthique algorithmique.

RGPD (Règlement Général sur la Protection des Données) :

- Il impose la minimisation des données utilisées, la finalité explicite des traitements, et le droit d'accès, de rectification ou de suppression pour les utilisateurs.
- Les données utilisées dans notre projet sont pseudonymisées (user_id, user_session), ce qui limite les risques de ré-identification.

IA Act (proposition européenne) :

- Ce règlement à venir prévoit des niveaux de risques selon les usages de l'IA. Notre projet serait classé en "IA à faible risque", mais il impose tout de même des obligations de transparence, de documentation et de contrôle humain.

Recommandations de la CNIL :

- La CNIL encourage l'utilisation de modèles explicables, l'audit des biais algorithmiques, et la mise en place de procédures d'analyse d'impact (PIA) pour certains traitements.

Mesures prises dans notre projet :

- Limitation des données à des logs pseudonymisés
- Visualisation des contributions des variables (feature importance, SHAP)
- Sensibilisation à l'explicabilité dans le choix des modèles
- Intégration possible d'une supervision humaine dans le processus de décision

- **CONCLUSION DE LA VEILLE**



L'environnement technologique, concurrentiel, sectoriel et réglementaire dans lequel s'inscrit notre projet confirme la pertinence de la démarche engagée. En mobilisant des modèles IA performants, explicables et respectueux du cadre juridique, notre solution s'inscrit pleinement dans les standards attendus aujourd'hui par les entreprises et les autorités. Cette veille continue permet également d'envisager des pistes d'amélioration futures, telles que l'intégration d'IA générative pour le contenu, ou la mise en œuvre de recommandations en temps réel via des architectures edge/cloud hybrides.

Indicateurs de performance (KPI) et plan de suivi

Le déploiement d'une solution d'intelligence artificielle n'a de valeur que s'il est mesurable dans sa capacité à produire des résultats concrets. Pour assurer le pilotage du projet et en démontrer l'impact, un ensemble d'indicateurs clés de performance (Key Performance Indicators – KPI) a été défini. Ces indicateurs couvrent à la fois les objectifs marketing, commerciaux et opérationnels liés à l'implémentation de la solution IA.

• OBJECTIFS DE SUIVI



L'objectif principal de cette phase est de permettre un suivi régulier, structuré et quantifiable de l'efficacité des modèles IA mis en place. Il s'agit d'observer l'évolution des résultats sur une période donnée et de comparer les performances avant et après l'intégration de l'intelligence artificielle dans les processus métiers.

Les indicateurs sélectionnés ont été choisis pour leur pertinence métier, leur mesurabilité dans les données disponibles et leur capacité à refléter directement l'impact de la personnalisation et de la prédiction.

• PLAN DE SUIVI ET GOUVERNANCE

Pour assurer un pilotage efficace, un plan de suivi structuré sera mis en œuvre dès la phase de déploiement :

Tableau de bord dynamique : un dashboard interactif (via Power BI ou Streamlit) permettra de suivre les KPI en temps réel. Il pourra être consulté par les équipes marketing, data et direction.

Points de contrôle réguliers :

- Un reporting hebdomadaire sur les KPI clés (taux de conversion, panier moyen, rebond)
- Un bilan mensuel incluant des comparatifs avant/après IA
- Une analyse trimestrielle des segments de clientèle et des performances des modèles

Mise à jour des modèles :

- Les modèles seront réentraînés périodiquement à partir de nouvelles données (approche continue)
- Des mécanismes de surveillance automatique (monitoring des dérives de performance) seront envisagés en phase d'industrialisation

Boucle de feedback métier :

- Les retours des équipes marketing seront intégrés pour ajuster les actions proposées par les modèles (recommandations, ciblage)
- Un processus d'amélioration continue sera mis en place pour aligner les résultats avec les objectifs business

KPI retenus

KPI	DÉFINITION	OBJECTIF VISÉ	FRÉQUENCE DE SUIVI
Taux de conversion	Pourcentage de visiteurs ayant effectué un achat	+10 % grâce à une meilleure personnalisation	Hebdomadaire
Panier moyen	Valeur moyenne d'un panier client	+8 % par effet de vente croisée (recommandation IA)	Hebdomadaire
Taux de rebond	Proportion d'utilisateurs quittant le site sans interaction	Réduction de 20 % par des suggestions plus engageantes	Hebdomadaire
Taux d'ouverture des emails ciblés	% d'e-mails ouverts parmi ceux envoyés avec contenu personnalisé	Augmenter l'efficacité des campagnes (objectif : +15 %)	Par campagne
ROI marketing	Retour sur investissement des actions marketing (CA généré / coût)	+15 % via un ciblage plus pertinent	Mensuel
Répartition par segments clients	Analyse des ventes par segment RFM ou cluster K-Means	Mieux piloter les actions spécifiques à chaque profil	Mensuel
Taux d'erreur de prévision (MAE / RMSE)	Écart entre prévisions de ventes et ventes réelles	Minimiser les erreurs pour optimiser stock & planning	Après chaque cycle de prévision

La mise en place de ces indicateurs et d'un plan de suivi rigoureux garantit une mesure précise de l'impact de la solution IA sur les ventes, l'engagement client et l'efficacité des campagnes marketing. Cette approche orientée résultats permet non seulement de valoriser les apports de l'IA auprès des décideurs, mais aussi d'ajuster dynamiquement les leviers d'action selon les données observées, dans une logique d'optimisation continue.



Étude des outils et technologies retenus

La sélection des outils constitue un levier stratégique essentiel dans la réussite du projet IA. Elle a été guidée par quatre critères principaux :

- la compatibilité avec les données disponibles,
- la pertinence vis-à-vis des objectifs du projet (recommandation, segmentation, prédiction),
- la facilité d'intégration dans l'environnement technique existant,
- la capacité à industrialiser la solution en cas de généralisation.

Nous avons structuré cette étude autour des différentes phases du cycle de vie du projet IA : acquisition et traitement des données, modélisation, visualisation, déploiement, et monitoring.

• PRÉPARATION ET TRAITEMENT DES DONNÉES

Python (langage principal)

- Pourquoi ce choix ?

Python est aujourd'hui le standard dans le domaine de la data science. Il propose une large écosystème de bibliothèques scientifiques et une grande communauté active, ce qui facilite le développement, le partage et la reproductibilité des analyses.

- Alternatives écartées : R (très bon pour la statistique, mais moins adapté aux workflows de machine learning à grande échelle), Java (plus performant mais beaucoup moins souple pour le prototypage IA).

• BIBLIOTHÈQUES PYTHON UTILISÉES :

BIBLIOTHÈQUE	RÔLE	POURQUOI CE CHOIX ?
pandas	Manipulation tabulaire des données	Manipulation puissante de DataFrames, très lisible
numpy	Calculs numériques	Utilisé comme base de nombreuses autres bibliothèques
gzip / csv	Lecture des fichiers compressés	Indispensable au traitement des .csv.gz

• MODÉLISATION IA

Scikit-learn

- C'est la bibliothèque la plus complète et stable pour des modèles classiques (K-Means, Random Forest, régressions...), avec des interfaces simples et standardisées.
- Avantage clé : permet de tester rapidement plusieurs algorithmes et pipelines de preprocessing.

XGBoost

- Très puissant pour les données tabulaires.
- Résistant au surapprentissage, performant en compétition Kaggle.
- Meilleure précision que les modèles linéaires ou les arbres simples.

Random Forest

- Complémentaire à XGBoost pour comparer les performances.
- Plus simple à entraîner, facile à interpréter.
- Utile en première approche avant tuning approfondi.

K-Means (clustering)

- Modèle standard pour les segmentations client RFM.
- Très rapide, lisible, actionnable pour les équipes marketing.

LightFM (recommandation hybride)

- Combine filtrage collaboratif et contenu produit (prix, catégorie...).
- Supporte bien le problème du cold-start.
- Plus simple à intégrer qu'un système de deep learning.

Modèles non retenus (et pourquoi) :

- Deep learning (RNN, LSTM) : trop lourd pour le cadre du projet, besoin de GPU, difficile à interpréter.
- Autoencoders pour recommandation : moins explicables, peu adaptés aux données disponibles.
- ARIMA/SARIMA : demandent une stationnarité stricte, difficile à appliquer sur des ventes e-commerce irrégulières.



Streamlit

- Crée facilement des dashboards interactifs.
- Intégrable avec les notebooks.
- Idéal pour exposer les résultats à des non-techniciens.

Power BI (option possible si infrastructure Microsoft)

- Ciblé business.
- Permet des connexions aux données en temps réel.

Matplotlib / Seaborn

- Analyse exploratoire des données.
- Visualisation claire des relations, tendances et clusters.

FastAPI

- Framework rapide et léger pour exposer des modèles IA sous forme d'API REST.
- Facilement déployable avec Docker ou via un cloud provider.

Docker

- Crée des environnements reproductibles et isolés.
- Permet de déployer la même version de code dans tous les environnements (local, cloud, production).

GitHub / GitHub Actions

- Versionning du code source, gestion collaborative.
- GitHub Actions permet l'automatisation des tests et du déploiement.

Prometheus + Grafana (prévu en cas d'industrialisation)

- Monitoring des modèles en production (temps de réponse, erreurs, dérive des performances).



Roadmap du projet

Le projet a été conçu selon une logique de phases progressives, chacune avec des objectifs, des livrables et des validations internes. La méthodologie employée repose sur un cycle de type agile, permettant des itérations courtes et des ajustements fréquents.

PHASE	DURÉE ESTIMÉE	OBJECTIFS PRINCIPAUX	LIVRABLES
1. Cadrage et audit	Sem. 1-2	Comprendre le besoin, auditer la qualité et structure des données	Note de cadrage, rapport d'audit
2. Préparation des données	Sem. 3-4	Nettoyage, agrégation, transformation des logs utilisateurs	Dataset prêt pour modélisation
3. Exploration et choix des modèles	Sem. 5	Comparaison d'algorithmes, sélection des meilleurs candidats	Notebook comparatif, métriques
4. Modélisation IA	Sem. 6-7	Entraînement des modèles : LightFM, KMeans, XGBoost, etc.	Modèles sauvegardés, code reproductible
5. Tests métier et ajustement	Sem. 8-9	Évaluer les résultats avec les objectifs business	Résultats KPI, ajustements, A/B tests
6. Restitution & visualisation	Sem. 10-11	Création de tableaux de bord, explication des recommandations	Dashboard, visualisation interactive
7. Suivi & plan d'amélioration continue	Sem. 12	Plan de monitoring, recommandations pour industrialisation	Plan de maintenance, fiche modèle



Conformité RGPD et gouvernance des données

La gouvernance des données est au cœur de la stratégie IA d'Amazing. Même si les données sont pseudonymisées, l'analyse de comportements utilisateurs impose une vigilance éthique et réglementaire à toutes les étapes.

• ANALYSE RGPD

PRINCIPE RGPD	APPLICATION DANS LE PROJET
Minimisation	Seules les colonnes strictement nécessaires sont utilisées (event_type, user_id, product_id, etc.)
Pseudonymisation	Les identifiants utilisateurs sont anonymes (aucun email, nom, localisation).
Finalité explicite	Les traitements sont destinés à la personnalisation marketing uniquement.
Transparence	Tous les traitements peuvent être documentés et expliqués.
Droits utilisateurs	Possibilité d'intégrer une interface permettant à l'utilisateur de refuser la personnalisation.

• ÉTHIQUE & IA

- **Biais algorithmiques** : les modèles sont testés sur différents segments pour détecter d'éventuelles discriminations.
- **Explicabilité** : importance des variables affichée pour tous les modèles via SHAP ou arbres de décision.
- **Contrôle humain** : les équipes marketing conservent le dernier mot sur les campagnes déclenchées.

• PERSPECTIVES D'INDUSTRIALISATION RGPD

Dans le cadre d'un déploiement réel, les actions suivantes sont recommandées :

- **Réalisation d'une Analyse d'Impact (PIA).**
- **Mise en place d'un registre de traitement IA.**
- **Désignation formelle d'un DPO référent pour le projet.**
- **Audit régulier des performances et des biais des modèles en production.**

Évaluation des moyens prévisionnels et estimation budgétaire

Afin de garantir la faisabilité opérationnelle du projet IA proposé à Amazing, il est indispensable d'évaluer les moyens humains, techniques et financiers à mobiliser. Cette estimation repose sur un déploiement de type preuve de concept (POC), sur une durée estimée de 12 semaines, avec l'objectif de tester, valider, puis itérer la solution avant une éventuelle industrialisation.

• MOYENS HUMAINS À MOBILISER

Le projet nécessitera une collaboration multidisciplinaire impliquant :

- **1 chef de projet IA** : supervision, planification, coordination avec les équipes métier.
- **1 data scientist** : responsable de l'analyse des données, du choix et de l'entraînement des modèles.
- **1 data engineer (ressource interne d'Amazing)** : mise à disposition des données depuis le Data Lake, traitement initial.
- **1 référent marketing** : interface métier, validation des résultats, tests A/B.
- **1 expert RGPD ou DPO** : validation de la conformité des traitements algorithmiques.

• MOYENS TECHNIQUES

Le projet s'appuie sur des outils principalement open-source et une infrastructure existante :

- **Environnement de développement** : Jupyter Notebook ou Google Colab.
- **Langage de programmation** : Python.
- **Bibliothèques** : scikit-learn, LightFM, XGBoost, Prophet, SHAP.
- **Visualisation** : Streamlit ou Power BI.
- **Déploiement envisagé** : Docker + FastAPI (si industrialisation).
- **Infrastructure** : Data Lake cloud d'Amazing (déjà en place).

• ESTIMATION BUDGÉTAIRE

POSTE	COÛT ESTIMÉ	DÉTAIL
Data scientist (3 mois)	~13 500 €	Analyse, modélisation, visualisation
Chef de projet (1 mois ETP)	~4 500 €	Coordination, gestion des livrables
Cloud / infrastructure	~1 200 €	Stockage, calcul, hébergement (POC uniquement)
Sensibilisation métier (1 journée)	~800 €	Ateliers d'appropriation des outils IA
Total POC (hors industrialisation)	~20 000 €	Montant prévisionnel complet

Note de faisabilité synthétique

Cette note vise à fournir une vue exécutive condensée, destinée aux décideurs ou sponsors du projet. Elle reprend les éléments clés permettant de valider la pertinence et la viabilité de la solution.

• CONTEXTE

- Amazing souhaite répondre à une baisse significative des ventes sur sa gamme "Basics", en particulier sur les produits de divertissement. Dans un environnement digital concurrentiel, l'entreprise s'oriente vers une exploitation plus fine de ses données client, via l'intelligence artificielle, pour personnaliser l'expérience utilisateur et optimiser ses actions marketing.

• OBJECTIFS DU PROJET

- Déployer une solution IA visant à :
 - Recommander des produits pertinents à chaque utilisateur,
 - Segmenter les clients selon leur comportement d'achat,
 - Anticiper les ventes pour ajuster l'offre et les campagnes.
- Offrir une visualisation claire et exploitable des résultats aux équipes métiers.

• SOLUTION PROPOSÉE

- **Recommandation personnalisée** : modèle LightFM (hybride).
- **Segmentation client** : analyse RFM + clustering K-Means.
- **Prévision des ventes** : modèles XGBoost, Prophet, Random Forest.
- Restitution des résultats via dashboards interactifs (Streamlit, Power BI).
- Données extraites du Data Lake cloud d’Amazing.

• ESTIMATION BUDGÉTAIRE

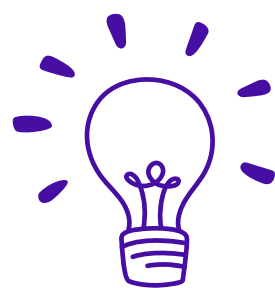
RISQUE	ACTION PRÉVUE
Adoption difficile côté métier	Ateliers de sensibilisation, dashboards visuels
Données incomplètes ou bruitées	Nettoyage, agrégation, robustesse des modèles
Biais algorithmiques	Tests multi-segments, interprétabilité (SHAP)
Risques RGPD	Données pseudonymisées, documentation, supervision DPO

• VALEURS ATTENDUE

- **+10 à 15 % de taux de conversion estimé via la personnalisation.**
- **Augmentation du panier moyen grâce à des recommandations ciblées.**
- **Meilleure anticipation des ventes pour ajuster le stock et les promotions.**
- **Outil d’aide à la décision basé sur des segments clients concrets.**

• CONCLUSION

La solution proposée est à la fois réaliste, responsable et à fort potentiel. Elle répond de manière ciblée aux besoins d’Amazing en conciliant performance marketing, compréhension client et conformité réglementaire. Le projet peut être lancé sous forme de POC avec un budget maîtrisé, et monté en puissance selon les résultats obtenus.



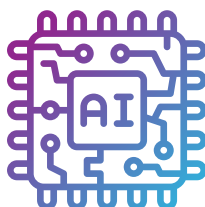
Conclusion



Ce projet a permis de démontrer la faisabilité et la pertinence d'une solution d'intelligence artificielle appliquée au secteur du e-commerce, dans un contexte où la performance commerciale dépend de plus en plus de la personnalisation et de l'analyse des données. En répondant à une problématique métier claire – relancer les ventes d'Amazing Basics – nous avons su proposer une approche fondée sur des modèles éprouvés, des outils accessibles, et une logique rigoureuse de cadrage, de test et de validation.



La combinaison des algorithmes retenus (LightFM, K-Means, XGBoost, Prophet, Random Forest) permet de couvrir l'ensemble du cycle de décision marketing : de la compréhension du client à la recommandation de produits, jusqu'à la prévision de la demande. Le projet intègre également des dimensions essentielles comme l'éthique algorithmique, la conformité RGPD, la visualisation des résultats, et l'accompagnement des équipes métiers. Ces éléments assurent une solution à la fois performante, responsable, et compréhensible pour tous les acteurs de l'entreprise.



En somme, cette mission illustre concrètement comment l'intelligence artificielle peut devenir un levier stratégique au service de la compétitivité et de l'expérience client. La méthodologie suivie, les outils choisis et les livrables proposés offrent une base solide pour un déploiement progressif à plus grande échelle, dans une logique d'amélioration continue. Ce projet est aujourd'hui prêt à passer de la phase d'expérimentation à une mise en œuvre opérationnelle, au service de la transformation digitale d'Amazing.