

Exigences du projet



Projet : Amazon Review Analysis

Auteur : Dyhia TOUAHRI

Date : 23 octobre 2025

Version 1.0

A propos de ce document

Ce document reprend les exigences, les normes et les conditions spécifiques auxquels le projet doit répondre pour être mené à bien. Le but de ce document est de comprendre les objectifs et les limites du projet ainsi que de s'assurer que tous les acteurs impliqués savent ce qu'on attend d'eux.

Les éléments repris dans ce document se base sur un prototype développé afin d'étudier la faisabilité et de donner un premier aperçu de la future solution qui sera développée par la suite dans le cadre de ce projet.

Table des matières

- 1. Introduction4
 - 1.1. Contexte du projet4
 - 1.2. Problématiques.....4
 - 1.3. L’objectif du projet4
 - 1.4. L’attendu du projet.....4
- 2. Périmètre5
 - 2.1. Inclusions5
 - 2.2. Exclusions5
 - 2.3. Besoins fonctionnels6
 - 2.4. Besoins non fonctionnels.....6
- 3. Contraintes.....6
- 4. Critères de succès :7

1. Introduction

1.1. Contexte du projet

Amazon, un des grands leaders du e-commerce, reçoit des milliers d'avis d'utilisateurs sur les produits vendus sur sa plateforme. Ces avis constituent une source précieuse pour les consommateurs dans leur processus de décision d'achat.

Toutefois, les produits peuvent avoir énormément d'avis et tous les avis n'ont pas la même valeur. Certains sont courts, certains sont biaisés et d'autres non pertinents. L'objectif de ce projet est de développer un système de scoring des commentaires pour tous les produits afin d'évaluer leurs pertinences et les mettre en avant.

1.2. Problématiques

- Tous les avis ne sont pas toujours pertinents.
- Les notes laissées par les consommateurs ne reflètent pas toujours la qualité réelle du produit.
- Amazon et les vendeurs ont besoin d'indicateur plus fiable basé sur la qualité des avis.
- Les commentaires les plus pertinents ne remontent pas forcément en priorité pour les consommateurs.

1.3. L'objectif du projet

Le projet vise à développer une solution intelligente capable d'analyser automatiquement les avis clients Amazon afin de déterminer les produits les plus pertinents selon un système de scoring.

La solution comprendra deux étapes principales : la catégorisation et la notation des avis selon un score de confiance, puis le calcul d'un score de pertinence global destiné à identifier et promouvoir les avis les plus pertinents pour chaque produit.

1.4. L'attendu du projet

- Classifier chaque avis des produits, en combinant plusieurs critères comme la pertinence du texte, sa longueur, la présence d'images, la confiance du modèle et la richesse en mots-clés.
- Mettre en avant les avis les plus utiles et fiables, afin que les utilisateurs puissent rapidement identifier les commentaires de qualité.
- Construire un tableau de bord pour visualiser les scores des avis, comparer leur pertinence et explorer les informations importantes.

2. Périmètre

2.1. Inclusions

Le projet couvre les activités suivantes :

- Collecte des données Amazon : utilisation d'un jeu de données d'avis produits Amazon stocké dans une base de données relationnelle.
- Nettoyage et préparation des données : traitement et harmonisation des données, notamment la suppression des doublons et la correction des incohérences.
- Catégorisation des avis : application d'un modèle d'intelligence artificielle pour classifier les avis avec un score de confiance associé. Les catégories identifiées :
 - "product quality or satisfaction" pour les commentaires qui parlent de la qualité des produits.
 - "product defect or damaged item" pour les avis qui parlent des produits défectueux ou endommagés.
 - "delivery issue or shipping delay" pour les avis qui dénoncent des problèmes de livraisons.
 - "customer service or support" pour les avis qui dénoncent des problèmes avec le service client.
- Calcul du poids des avis : définition et mise en œuvre d'une formule de pondération intégrant les critères suivants :
 - 30% taille de la description du produit (utilisation de la formule Gaussienne pour déterminer la taille).
 - 20% si la description contient une image.
 - 10% si l'utilisateur a passé une commande.
 - 15% si l'utilisateur a laissé un score de 1 ou 5 (extrémité).
 - 25% le score est calculé à l'aide d'un algorithme, qui évalue le sentiment du texte global. Plus la description exprime un sentiment positif ou négatif marqué, plus elle est considérée comme pertinente.
- Analyse et visualisation : création de graphiques représentant la distribution des notes et la classification des avis, ainsi que la comparaison des scores par catégorie ou type de produit.

2.2. Exclusions

Le projet ne comprend pas les éléments suivants :

- Collecte automatique via API : seules des données statiques, récupérées une fois, seront utilisées, en raison de contraintes légales et de limites d'accès.
- Interface utilisateur : aucun développement d'interface web ou de dashboard interactif n'est prévu.
- Intégration avec les systèmes réels Amazon.
- Actions automatiques sur la plateforme Amazon

2.3. Besoins fonctionnels

En se basant sur le prototype, le projet doit permettre de :

- Charger les données Amazon : charger le jeu de données contenant les avis produits dans un data lake depuis une base de données relationnelle.
- Traitement des données (ETL) : récupération à partir du data lake, traitement des doublons, de texte et jointures et anonymisation des données personnelles.
- Stockage des données : les données sont ensuite stockées dans un datawarehouse et une base de données NoSQL.
- Classification et calcul des poids pour les avis : appliquer un algorithme de classification puis une formule de pondération basée sur plusieurs critères. Ces deux éléments vont permettre d'évaluer la pertinence des avis. La pondération des variables pourrait être amenée à changer.
- Stocker les résultats : les résultats sont ensuite enregistrés dans le datawarehouse.
- Visualiser les résultats : des visualisations permettant d'analyser les résultats.

2.4. Besoins non fonctionnels

Le projet doit également respecter les exigences suivantes :

- Sécurité : aucune donnée personnelle n'est affichée ou exposée dans les jeux de données.
- Accessibilités : La solution doit pouvoir être exécutée sur n'importe quel environnement de travail.
- Maintenabilité : le code doit être modulable, lisible et correctement commenté.
- Scalabilité : la solution doit pouvoir évoluer pour intégrer, à l'avenir, des flux de données via API ou d'autres sources externes.

3. Contraintes

Le projet doit tenir en compte les contraintes suivantes :

- Temps limité : le projet est réalisé dans le cadre d'un projet académique.
- Multilinguisme des avis : les avis étant rédigés en plusieurs langues, il est important de choisir un algorithme capable de prédire correctement les catégories.
- Données : accès en lecture seule sur la base de données relationnelle.
- Limite des données : seules des données simulées ou publiques sont utilisées, sans accès à l'API Amazon.
- Ressource humaine : sur certains points l'équipe est encore junior sur les sujets (ML).
- Infrastructure : le traitement se fait sur une infrastructure locale, en privilégiant des solutions gratuites et open source.
- Données personnelles : dans le projet les données à caractères personnelles identifiées doivent être anonymisés.

4. Critères de succès :

Le projet sera considéré comme réussi si :

- Exécution sans erreur : la solution peut être exécutée correctement dans son environnement prévu.
- Pondération justifiée et paramétrable : le calcul du poids des avis est documenté et peut être reproduit.
- Résultats cohérents : les scores pondérés sont en corrélation avec les notes globales des avis, permettant une analyse fiable.
- Documentation : la solution complète est documentée.