**Rapport d'Analyse des Sentiments et Profiling des Audiences**

**1. Introduction**

**Objectif du projet**

Ce projet vise à analyser les commentaires des clients à partir du dataset **Amazon Review Polarity Dataset** afin d'extraire des insights pertinents pour la stratégie marketing. Nous avons utilisé des techniques de **traitement du langage naturel (NLP)** et de **clustering** pour mieux comprendre les segments d’utilisateurs et leur perception des produits.

**Description du dataset**

**Liens du dataset :**

[**https://www.kaggle.com/datasets/kritanjalijain/amazon-reviews**](https://www.kaggle.com/datasets/kritanjalijain/amazon-reviews)

Le **Amazon Review Polarity Dataset** contient des avis clients annotés en deux catégories principales :

* **Positif** (reviews avec des notes de 4 et 5)
* **Négatif** (reviews avec des notes de 1 et 2)

Chaque classe contient **1,800,000 échantillons d'entraînement** et **200,000 échantillons de test**. Pour accélérer l'analyse, nous avons extrait un échantillon représentatif de la base de données.

**2. Méthodologie**

**Étapes du pipeline d’analyse**

1. **Extraction des données** à partir du dataset.
2. **Prétraitement** : nettoyage du texte, tokenization, suppression des stopwords.
3. **Analyse de sentiments** avec la méthode **VADER**.
4. **Clustering** des avis pour identifier des groupes homogènes.
5. **Visualisation et interprétation des résultats**.

**Outils et bibliothèques utilisés**

* **NLTK** et **spaCy** pour le NLP.
* **Scikit-learn** pour le clustering (K-Means, PCA).
* **Seaborn** et **Matplotlib** pour les visualisations.

**3. Résultats et Visualisations**

**3.1 Analyse des sentiments**

La **répartition des sentiments** obtenue via **VADER** montre une large prédominance des avis **positifs** :

* **Avis positifs** : ~75%
* **Avis négatifs** : ~22%
* **Avis neutres** : ~3%

🔹 **Interprétation** : La majorité des clients expriment une satisfaction envers les produits, mais une part significative (~22%) reste insatisfaite, ce qui représente un levier d'amélioration.

**3.2 Clustering des avis**

Nous avons appliqué un algorithme de **K-Means** (k=2) pour segmenter les avis. Après une **réduction de dimension avec PCA**, les clusters sont visualisés :

* **Cluster 0** : Regroupe majoritairement des avis sur des produits **liés au divertissement (films, albums, musique)**.
* **Cluster 1** : Regroupe des avis principalement sur des **livres et auteurs**.

🔹 **Interprétation** : On distingue deux grandes catégories d’intérêts. Cette segmentation peut aider à personnaliser la stratégie marketing selon le type de produit.

**3.3 Word Clouds par cluster**

L'analyse des **termes les plus fréquents** dans chaque cluster donne les résultats suivants :

* **Cluster 0** : Mots dominants → "great", "album", "movie", "music"
* **Cluster 1** : Mots dominants → "book", "read", "author", "story"

🔹 **Interprétation** : Le **Cluster 0** est centré sur les médias et divertissements, tandis que le **Cluster 1** est axé sur les livres et la lecture.

**3.4 Correspondance entre Clusters et Sentiments**

Le **tableau de contingence** montre :

* **Cluster 0** : 3918 avis négatifs vs 3828 positifs (répartition équilibrée).
* **Cluster 1** : 1036 avis négatifs vs 1218 positifs.

🔹 **Interprétation** : Le Cluster 0 contient plus d'avis **mitigés**, alors que le Cluster 1 semble plus favorablement perçu.

**4. Insights et Recommandations Marketing**

**4.1 Identification de segments clients**

✅ **Segment "Médias & Divertissement" (Cluster 0)** :

* Un nombre élevé d'avis mitigés → Il pourrait être intéressant d'explorer les raisons des avis négatifs.
* La stratégie pourrait inclure **plus d'engagement client** et des **offres promotionnelles ciblées**.

✅ **Segment "Livres & Auteurs" (Cluster 1)** :

* Ce segment montre un **taux plus élevé d’avis positifs**.
* Stratégie recommandée : **Mise en avant des best-sellers et recommandations personnalisées**.

**4.2 Actions concrètes**

📌 **Améliorer le service client** :

* Identifier les problèmes récurrents (ex : retours négatifs liés à la livraison ou au produit).
* Réagir rapidement aux avis négatifs en proposant des solutions (remboursements, assistance rapide).

📌 **Personnalisation du marketing** :

* Cibler les utilisateurs avec des offres spécifiques en fonction de leur cluster.
* Utiliser les insights pour adapter la communication et optimiser l'expérience utilisateur.

📌 **Monitoring continu** :

* Mettre en place un **dashboard de suivi des avis** pour suivre l'évolution des sentiments en temps réel.
* Expérimenter d'autres **modèles d'analyse NLP** (BERT, Transformers) pour améliorer la précision.

**5. Conclusion**

🎯 **Synthèse des analyses** :

* La majorité des avis clients sont **positifs**, mais un segment non négligeable exprime des **insatisfactions**.
* Le **clustering** a révélé deux profils d’utilisateurs distincts (**Médias & Divertissement vs Livres & Auteurs**).
* La **répartition des sentiments** au sein de ces clusters montre des différences significatives, influençant la perception des produits.

🚀 **Perspectives d’amélioration** :

* **Tester d'autres techniques de NLP** comme BERT pour améliorer la classification des sentiments.
* **Intégrer des données supplémentaires** (comportements d’achat, fréquence d’interaction) pour enrichir l’analyse.
* **Automatiser l’analyse des avis** en continu pour une meilleure réactivité aux tendances clients.

✅ **Impact pour l'entreprise** : Ce projet démontre comment une **analyse avancée des avis clients** permet d’optimiser la stratégie marketing et d’améliorer l’expérience utilisateur.

📢 **Prochaines étapes** :

* Déploiement d’un **modèle en production** pour suivre les tendances des avis en temps réel.
* Mise en place d’un **système de recommandations personnalisé** basé sur le sentiment des utilisateurs.

🔍 **Résumé en une phrase** : Grâce à l’analyse des sentiments et au clustering des avis Amazon, nous avons pu extraire des insights marketing actionnables pour optimiser l’engagement et la satisfaction des clients.