**Analyse des Corrélations et Associations**

**Objectif :** Identifier les relations entre sentiments des clients, produits mentionnés et problèmes récurrents.

**1. Extraction des Produits/Marques Mentionnés**

* **Modèle utilisé** : **dslim/bert-base-NER** (Reconnaissance d'entités nommées).
* **Méthode** : Identifier les noms de produits ou de marques dans la colonne**reviewText**.
* **Exemple Review** :

Si un avis contient : *"J'adore le shampoing Bvlgari, il sent bon !"*,

* **Exemple de sortie** :
  + **Produits/Marques extraits** : Bvlgari.
  + **Fréquence** : Nombre de fois où chaque produit/marque est mentionné.

**2. Analyse des Corrélations entre Sentiments et Produits/Marques**

* **Modèle utilisé** : **distilbert-base-uncased-finetuned-sst-2-english** (Analyse de sentiment).
* **Méthode** : Croiser les sentiments (sentiment) avec les produits/marques extraits.
* **Exemple de sortie** :
  + **Bvlgari** : 80% des avis positifs, 20% des avis négatifs.
  + **Aqua Velva** : 60% des avis positifs, 40% des avis négatifs.
  + **Autre Marque X** : 70% des avis positifs, 30% des avis négatifs.

**3. Identification des Problèmes Récurrents**

* **Modèle utilisé** : **sentence-transformers/all-MiniLM-L6-v2** (Clustering et extraction de mots-clés).
* **Méthode** : Analyser les mots-clés associés aux avis négatifs (label = 1) dans reviewText.
* **Exemple de sortie** :
  + **Problèmes récurrents** : "odeur désagréable", "qualité médiocre", "service client défaillant".
  + **Fréquence** : Pourcentage d'avis négatifs mentionnant chaque problème.

**Exemple de Sortie Adapté**

1. **Produits/Marques Extraits et Sentiments Associés**

|  |  |  |  |
| --- | --- | --- | --- |
| Produit/Marque | Avis Positifs (%) | Avis Négatifs (%) | Problèmes Récurrents |
| Bvlgari | 80% | 20% | Aucun problème majeur |
| Aqua Velva | 60% | 40% | Odeur désagréable |
| Autre Marque X | 70% | 30% | Qualité médiocre |

1. **Problèmes Récurrents Identifiés**

|  |  |
| --- | --- |
| Problème | Fréquence dans les Avis Négatifs (%) |
| Odeur désagréable | 40% |
| Qualité médiocre | 35% |
| Service client défaillant | 25% |

1. **Modèles Utilisés à Chaque Étape**

|  |  |  |
| --- | --- | --- |
| Étape | Modèle Utilisé | Résultat |
| Extraction des Produits/Marques | dslim/bert-base-NER | Liste des produits/marques mentionnés dans les avis. |
| Analyse des Sentiments | distilbert-base-uncased-finetuned-sst-2-english | Corrélations entre les sentiments et les produits/marques. |
| Identification des Problèmes | sentence-transformers/all-MiniLM-L6-v2 | Liste des problèmes récurrents et leur fréquence dans les avis négatifs. |

|  |  |
| --- | --- |
| Feature | Associated Model(s) |
| Trend detection | sentence-transformers/all-MiniLM-L6-v2 (semantic similarity) |
| Sentiment analysis | distilbert-base-uncased-finetuned-sst-2-english |
| Named entity recognition | dslim/bert-base-NER |
| Topic classification | facebook/bart-large-cnn (text summarization for analysis) |
| Correlation & association analysis | Statistical models (PCA, K-Means, Bayesian networks) |