

DÉCOUVREZ COMMENT APPLIQUER UNE ANALYSE NLP "EXPRESS" AUX AVIS CLIENTS



Dans le monde dynamique des affaires, une analyse rapide des retours clients par rapport à une marque ou un produit distinct est devenue un requis, surtout lorsqu'elle permet de répondre à des idées préconçues venant des experts métiers ou des décideurs d'entreprise. Découvrir ainsi le ressenti des clients et ce qui motive une bonne ou une mauvaise note ou encore être capable d'identifier les leviers d'amélioration de cette relation devient une stratégie essentielle pour éclairer la prise de décision.

Méthodologie

Vous êtes généralement confrontés à deux types de problématiques :

- Une problématique large nécessitant une phase exploratoire poussée des avis.
Par exemple : Qu'est ce qui explique une bonne ou une mauvaise note client ? Ou comment améliorer notre CSAT ?
- Une problématique spécifique.
Par exemple : Est-ce que nos clients sont satisfaits d'un produit/service en particulier ? Si non quelles sont les caractéristiques qui déplaisent et comment les améliorer ?

CAS D'UNE PROBLÉMATIQUE LARGE DU TYPE « COMMENT EXPLIQUER L'INSATISFACTION GLOBALE DES CLIENTS ? »

1 Compréhension du besoin et étude de la faisabilité :

- L'objectif de l'analyse doit être **clairement défini avec les métiers**, ce qui permettra de faire rapidement le bon choix des méthodes.
- Le nombre d'avis recueillis sur la période d'analyse doit être **significatif**. Le taux de retours doit être comparable à celui des périodes précédentes ou à celui validé lors d'une étude similaire.
- La diversité des avis ainsi que leur qualité sémantique sont des facteurs déterminants pour décider de la faisabilité de l'analyse. Et dans ce type d'analyse, **la qualité prévaut sur la quantité**. Voici quelques exemples d'avis client de mauvaise qualité sémantique : « Service client OK » ou « J'aime ce produit ».
- **Les ressources informatiques doivent être adaptées à la tâche à réaliser** surtout lorsque vous disposez d'une grande quantité de données ou vous envisagez d'appliquer des modèles/tâches complexes (ex : traduction d'un grand nombre d'avis, modélisation de topics, etc.)

2 Préparation des avis clients :

- **Réduction du spectre d'analyse** dans la mesure du possible. Selon le contexte, se focaliser sur les avis négatifs ou positifs : le plus simple serait d'utiliser la note attribuée par le client pour catégoriser les avis en positifs-neutres-négatifs. Il est possible aussi - si une question standard NPS a été posée dans le questionnaire de satisfaction - d'utiliser la définition interne de l'entreprise ou la définition générique **pour identifier les clients promoteurs et détracteurs**. A défaut, il est nécessaire de faire une analyse du sentiment. Certaines entreprises utilisent des outils de gestion et de recueil des avis clients ; tels que *TrustYou* ou *trustpilot* ; qui offrent la possibilité de catégoriser les avis par thématique ou/et par sentiment et vous feront donc gagner un temps précieux pour la suite de l'analyse.
- **Normalisation du texte** : mettre tous les caractères en minuscules (sinon « nlp » et « NLP » seront considérés comme deux mots distincts), normalisation Unicode, suppression des caractères spéciaux, gérer les caractères répétitifs (exemple : « Bonjouuuur » pour « bonjour »), etc.
- **Tokenization** : consiste en la transformation du texte en une séquence de tokens. Cette représentation facilite les tâches d'extraction ou de filtrage d'information selon le contexte.
- **Gestion des « stopwords »** : Ce sont des mots qui n'apportent généralement pas beaucoup de sens à votre analyse. En Python, les bibliothèques NLTK ou spaCy sont recommandées pour cette tâche. Ces dernières contiennent des modules pour vous donner la liste des stopwords dans plusieurs langues et pour les filtrer de votre texte.

3 Exploration des avis

- **Utilisation de la méthode du POS-Tagging (Part-of-Speech tagging) pour filtrer les tokens**. Vous pouvez vous focaliser sur les tokens ayant une des fonctions grammaticales suivantes : **Nom, Adjectif, Adverbe** en préparation de l'input d'un WordCloud. **Le POS-Tagging est une tâche qui consiste à attribuer à chaque mot un « tag » grammatical** comme nom, verbe, déterminant, etc.
- **Lemmatisation ou normalisation des mots** : Réduire les tokens à leurs formes de base (forme canonique) et réduire ainsi la dimension du vocabulaire.
- **Création d'un WordCloud** : Les termes les plus visibles dans le nuage de mots sont les plus fréquents dans le corpus. Cela peut paraître simple et sans intérêt, mais au contraire, un WordCloud permet d'avoir une vue d'ensemble du contenu des avis et d'identifier la nécessité d'un prétraitement supplémentaire des données.

4 Extraction des thématiques

- Une des méthodes qui pourraient être utilisées est le **topic modeling** (apprentissage non-supervisé) permettant d'extraire les thématiques principales d'un corpus. Un des modèles les plus utilisés est le **LDA (Latent Dirichlet Allocation)**. Le modèle LDA est un **modèle probabiliste génératif qui suppose que chaque document (avis) est généré à partir d'une combinaison de thèmes et que chaque thème est une distribution de mots**.
- L'input de ce modèle est une **représentation numérique du texte** (avis ou parties de l'avis). **Après «cleaning»** les données seront stockées dans un dictionnaire et puis converties en « **Bag of words** » (mot-occurrence).
- L'entraînement du modèle est un processus itératif dans lequel **il est important de s'assurer du bon nombre de thèmes**. Ceci nécessite donc une analyse des résultats (nous recommandons d'utiliser le package *pyLDAvis* en Python) en essayant dans un premier temps de comprendre les topics en listant les termes les plus fréquents et représentatifs de chaque topic, et dans un second temps de mesurer la distance entre les topics ainsi que leur cohérence en évaluant le degré de similitude sémantique entre les mots qui les représentent le mieux.

5 Aller plus loin : Description des thématiques

- Appliquer le modèle sur les avis ou des parties des avis **pour les catégoriser par thématique (un avis peut appartenir à plusieurs catégories)** puis extraire les chunks noms/adjectifs de chaque topic en utilisant les patterns grammaticaux pour mieux décrire les thèmes.

CAS D'UNE PROBLÉMATIQUE SPÉCIFIQUE DU TYPE « QUELLES CARACTÉRISTIQUES PRODUITS SONT À AMÉLIORER ? »

1 Compréhension du besoin et étude de la faisabilité

- Reprise des étapes du 1er cas
- Connaissant la thématique principale, vérifier en utilisant un dictionnaire de mots-clés s'il y a suffisamment d'avis qui traitent du sujet.

2 Préparation des avis clients :

- Filtrer sur les avis qui se rapportent à la thématique étudiée. Selon le niveau de technicité du sujet, un appel aux experts métiers peut s'avérer indispensable **pour créer ou valider un dictionnaire de mots-clés ou d'expressions liés à la thématique** en question.
- **Eviter de filtrer directement sur la note globale** (pour séparer avis positif/négatif) car un avis peut traiter de plusieurs thématiques et certaines peuvent être considérées moins critiques que d'autres pour le client.
- **Cleaning et mise en forme** (Normalisation, Tokenization, gestion des stopwords, etc.)

3 Exploration des avis avec un WordCloud

- Reprise des étapes du 1er cas
- Déduction des thématiques abordées (Par exemple : la qualité, le prix, le SAV, etc.).

4 Description des thématiques (caractéristiques)

- **Création d'un dictionnaire pour chaque thématique** utilisant les mots les plus fréquents dans le corpus (vous pouvez utiliser la sortie du **WordCloud** pour créer ces dictionnaires ou se référer à une expertise métier)
- **Mapping des dictionnaires dans les avis** : permet d'évaluer l'importance ou le poids des caractéristiques et leur criticité pour les clients.
- Utilisation des patterns grammaticaux pour extraire les chunks nom/adjectif rattachés à chaque topic

Principales difficultés et solutions

Une des principales difficultés est **la représentativité de l'échantillon d'avis à analyser**. En effet, les avis qui sont recueillis ne sont pas tous exploitables : contrainte de récence (période d'analyse), de qualité, de contexte (Pour l'analyse de la satisfaction client par rapport à un produit A, les avis qui parlent que du produit B ne seront pas retenus), etc. **Il est donc important de s'assurer, tout filtre appliqué, que vous n'êtes pas en train d'analyser un évènement rare** et dans ce cas penser à utiliser des modèles pré-entraînés sur des corpus volumineux.

Ensuite, **il y a la nécessité d'une expertise métier pour justifier certains choix analytiques**. Par exemple, les thèmes résultants d'un modèle **LDA** peuvent ne pas être clairs sémantiquement et il est donc difficile de leur attribuer une catégorie.

Il est recommandé d'impliquer des experts métier dans l'étape d'analyse et d'interprétation, ou encore dans le choix des paramètres d'entraînement des modèles comme le nombre de thèmes d'un LDA.

Et enfin, certaines tâches de NLP peuvent nécessiter une puissance de calcul significative et une gestion optimisée de la mémoire des machines. Une des solutions est d'**essayer de quantifier le coût en ressources de chaque tâche et penser à diviser les traitements en « batch »**. Dans un environnement cloud, **privilégier les modules embarqués aux ressources et librairies externes**.

Par exemple : un google Sheet contient des fonctions très efficaces de détection de la langue d'un texte et de traduction vers une autre langue.

Bénéfices

Mettre en place une méthodologie simple et claire d'analyse des avis client permet :

- > D'anticiper les demandes en automatisant certaines tâches comme le recensement préalable des thématiques abordées par les clients et catégoriser les avis.
- > Améliorer l'efficacité opérationnelle
- > Prise de décision rapide et amélioration/personnalisation de l'expérience client

Par exemple : la détection rapide d'une anomalie ou un défaut de fabrication d'un produit suite à un retour client.



Si vous souhaitez approfondir le sujet NLP ou en savoir plus sur l'expertise de notre Data Scientist, Bechir, n'hésitez pas à nous contacter : contact@aid.fr