

Promotion 2025, 5ème année, cycle ingénieur

Majeure Data & Intelligence Artificielle

Graph & Mining



WEBCRAPING PRE-PROJECT : CARTE INTERACTIVE DE LA SATISFACTION DANS L'HOTELLERIE ET LA RESTAURATION

Par Romain GIRAUD, Lily LA MORELLA, Christelle LE VAN KHANH



Date : vendredi 13 décembre 2024

TABLE DES MATIERES

I : Introduction 3

II : Objectifs du projet..... 4

III : Méthodologie..... 5

 1.1 Collecte des données (Web Scraping)..... 5

 1.2 Analyse des données (NLP) 5

 1.3 Création de la carte interactive..... 6

 1.4 Rapport de synthèse 6

IV : Technologies et outils et résultats attendus..... 7

V : Conclusion..... 8

I : INTRODUCTION

Dans un monde de plus en plus connecté, les avis en ligne jouent un rôle crucial dans la perception des services proposés par les hôtels et les restaurants. Ces avis sont une source précieuse d'informations pour comprendre les attentes des clients, identifier les tendances et améliorer les services. Ce projet vise à développer une carte interactive qui présente les niveaux de satisfaction des clients dans l'hôtellerie et la restauration, en fonction des avis collectés sur des plateformes spécialisées comme Booking.com, TripAdvisor, Yelp, ou encore Google Reviews.

Cette carte permettra non seulement de visualiser la satisfaction régionale dans ces deux secteurs, mais également d'analyser les avis grâce à des techniques de traitement du langage naturel (NLP). En combinant le web scraping, l'analyse de texte, et la visualisation de données, ce projet offre une solution innovante pour explorer les expériences clients et fournir des recommandations concrètes aux acteurs concernés.

II : OBJECTIFS DU PROJET

L'objectif principal de ce projet est de développer une carte interactive qui affiche, par région, le niveau de satisfaction des clients dans les secteurs de l'hôtellerie et de la restauration. Pour atteindre cet objectif, plusieurs sous-objectifs seront poursuivis :

- Récupérer des données d'avis clients sur des plateformes en ligne, en utilisant des techniques de web scraping efficaces.
- Analyser les données textuelles pour identifier les thèmes récurrents et évaluer le sentiment général des avis.
- Proposer une visualisation intuitive et interactive des résultats, permettant aux utilisateurs d'explorer les tendances régionales.
- Fournir des recommandations aux hôteliers et restaurateurs pour améliorer leurs services et mieux répondre aux attentes des clients.

En combinant ces approches, le projet vise à offrir une solution complète pour analyser et comprendre la satisfaction client dans ces deux secteurs.

III : MÉTHODOLOGIE

1.1 COLLECTE DES DONNEES (WEB SCRAPING)

La première étape du projet consiste à collecter des données à partir de plateformes d'avis en ligne. Les deux secteurs ciblés, hôtellerie et restauration, seront explorés à travers plusieurs sources principales. Pour les hôtels, des plateformes comme Booking.com, Expedia, et TripAdvisor seront utilisées. Pour les restaurants, les avis seront extraits de Yelp, Google Reviews, et TripAdvisor.

Afin d'extraire ces données, des outils de scraping tels que BeautifulSoup ou Scrapy seront employés. Ces bibliothèques permettent de parcourir les pages web et d'extraire des éléments spécifiques, comme les notes globales, le texte des avis, et la localisation des établissements. Dans le cas où certaines plateformes utilisent des contenus dynamiques générés par JavaScript, des outils comme Selenium pourront être utilisés pour interagir avec ces pages et extraire les données nécessaires.

Les données collectées seront ensuite structurées et sauvegardées dans des formats tels que CSV ou JSON pour une manipulation ultérieure. Une base de données relationnelle, comme SQLite ou PostgreSQL, sera mise en place pour centraliser les informations. Cette base facilitera le stockage, la manipulation et l'analyse des données.

1.2 ANALYSE DES DONNEES (NLP)

Une fois les données collectées, une analyse approfondie sera réalisée en utilisant des techniques de traitement du langage naturel (NLP). Cette analyse comprendra plusieurs étapes. Tout d'abord, les données textuelles seront nettoyées pour supprimer les éléments inutiles, comme les doublons, les fautes d'orthographe, et les mots vides (stop words). Ce prétraitement est crucial pour garantir la qualité des résultats.

Ensuite, des algorithmes de NLP, comme LDA (Latent Dirichlet Allocation), seront utilisés pour extraire les thèmes principaux des avis. Ces thèmes permettront d'identifier les éléments les plus fréquemment mentionnés par les clients, tels que la propreté, le service, ou encore la qualité de la nourriture. Parallèlement, une analyse de sentiment sera réalisée pour classer les avis en positifs, négatifs, ou neutres. Des outils comme VADER, TextBlob, ou spaCy seront utilisés pour cette tâche.

Enfin, un modèle statistique sera développé pour évaluer l'impact de chaque facteur (par exemple, la propreté ou le service) sur la satisfaction globale. Cette analyse permettra de quantifier la contribution de chaque critère à l'expérience client.

1.3 CREATION DE LA CARTE INTERACTIVE

Une fois les données analysées, les résultats seront intégrés dans une carte interactive, qui constituera le produit final du projet. Cette carte sera développée à l'aide de bibliothèques de visualisation comme Folium ou Plotly. Elle affichera les niveaux de satisfaction des hôtels et restaurants par région, en utilisant un code couleur pour représenter les différents scores. Par exemple, des zones avec une forte satisfaction seront marquées en vert, tandis que celles avec une satisfaction faible seront en rouge.

La carte sera enrichie de fonctionnalités interactives pour améliorer l'expérience utilisateur. En passant la souris sur une région ou un établissement, des info-bulles afficheront des détails tels que le nombre d'avis, les thèmes récurrents identifiés, ou encore les notes moyennes. Des filtres permettront également de sélectionner les données à afficher, par exemple en fonction du secteur (hôtellerie ou restauration) ou de la région.

1.4 RAPPORT DE SYNTHESE

Enfin, un rapport complet sera rédigé pour présenter les résultats du projet. Ce rapport analysera les tendances régionales en matière de satisfaction dans l'hôtellerie et la restauration, en mettant en évidence les points forts et les axes d'amélioration pour chaque région. Les conclusions seront accompagnées de recommandations concrètes pour aider les professionnels de ces secteurs à mieux répondre aux attentes des clients.

Ce rapport sera complété par une présentation interactive ou un site web, qui permettra de partager les résultats avec un public plus large. Ces supports incluront des visualisations claires et des graphiques pour illustrer les résultats de manière intuitive.

IV : TECHNOLOGIES ET OUTILS ET RÉSULTATS ATTENDUS

Le projet s'appuiera sur plusieurs technologies pour atteindre ses objectifs :

- Langages de programmation : Python sera utilisé pour toutes les étapes, de la collecte des données à l'analyse et la visualisation.
- Bibliothèques Python :
 - Pour le scraping : API, BeautifulSoup, Scrapy, Selenium.
 - Pour le NLP : spaCy, NLTK, TextBlob.
 - Pour la visualisation : Folium, Plotly, Dash.
 - Pour le stockage et la manipulation des données : Pandas, SQLite, PostgreSQL.

Le projet aboutira à plusieurs livrables concrets. Tout d'abord, une base de données structurée sera créée, contenant les avis clients avec leurs notes, leur localisation, et les thèmes associés. Ensuite, une analyse détaillée des sentiments et des thèmes récurrents sera réalisée. Ces résultats seront intégrés dans une carte interactive, qui permettra de visualiser les niveaux de satisfaction par région.

Enfin, un rapport de synthèse et une présentation interactive fourniront une analyse approfondie des résultats, accompagnée de recommandations exploitables pour améliorer la satisfaction client dans les secteurs de l'hôtellerie et de la restauration.

V : CONCLUSION

En combinant des techniques de scraping web, de traitement du langage naturel, et de visualisation interactive, ce projet offre une solution innovante pour explorer la satisfaction des clients dans les secteurs de l'hôtellerie et de la restauration. Il permettra non seulement de répondre aux critères académiques, mais aussi de fournir des insights précieux aux professionnels de ces secteurs. En respectant les étapes décrites et les technologies proposées, ce projet aboutira à des résultats concrets et utiles, tant sur le plan technique que pratique.