

Projet Intégré : Évaluation de la Qualité et de la Perception des Restaurants

Objectif du Projet Ce projet vise à combiner les données de conformité sanitaire fournies par l'API Alim'confiance avec l'analyse des sentiments des avis clients pour donner une vue d'ensemble de la qualité des établissements de restauration. Cette analyse globale permettrait d'identifier des corrélations entre les évaluations sanitaires officielles et la perception des consommateurs, offrant aux parties prenantes une vision complète pour améliorer la sécurité et la satisfaction client.

Technologies Utilisées - Python : Extraction et traitement des données provenant de l'API Alim'confiance et des avis clients. - Confluent Kafka : Ingestion des données en temps réel, permettant un traitement continu. - Apache Spark : Transformation des données, analyse de sentiments, et préparation des données .
Docker : Conteneurisation de tous les services pour faciliter le déploiement et la portabilité. - Elasticsearch : Stockage et indexation des données pour des recherches rapides et efficaces. - Kibana : Visualisation des données, création de tableaux de bord et visualisation géographique. - OpenAI API : Analyse des sentiments pour les avis clients.

Architecture du Projet

Workflow et Schéma d'Architecture:

1. Ingestion des Données de Contrôles Sanitaires (API Alim'confiance) :

Les informations sur les établissements et leurs contrôles sanitaires sont extraites de l'API Alim'confiance et envoyées dans Kafka.

2. Ingestion des Avis Clients :

Les avis clients sont extraits à partir de sources disponibles (plateformes d'avis ou réseaux sociaux, en respectant les règles d'utilisation).

Les avis sont également envoyés à Kafka pour une ingestion en flux continu.

3. Traitement des Données :

Analyse des Sentiments : Utilisation de l'API OpenAI pour identifier le sentiment général (positif, neutre, négatif) et les sentiments par aspect (qualité de la nourriture, service, ambiance).

Transformation des Données : Spark nettoie et enrichit les données, en ajoutant des informations telles que la date, la région, le département et le type d'établissement.

4. Indexation et Stockage :

Les données enrichies sont stockées dans Elasticsearch, indexées par établissement, par date, par région, et par catégorie (avis clients et contrôles sanitaires).

5. Visualisation et Analyse :

Kibana est utilisé pour créer des tableaux de bord interactifs, qui permettent de suivre la conformité sanitaire et l'expérience client en temps réel.

Fonctionnalités du Projet

- 1. Suivi des Contrôles Sanitaires:

Objectif : Fournir une vue d'ensemble des niveaux d'hygiène pour chaque établissement.

- 2. Analyse des Sentiments des Avis Clients:

Objectif : Quantifier le sentiment des clients pour chaque établissement en fonction de l'expérience (service, nourriture, ambiance).

- 3. Corrélation entre Sentiment Client et Hygiène Sanitaire:

Objectif : Identifier les corrélations entre la satisfaction des clients et le niveau de conformité

sanitaire.

- 4. Analyse Géographique et Temporelle:

Objectif : Identifier les zones géographiques avec des tendances sanitaires ou de satisfaction particulières.

- 5. Rapports et Indicateurs de Qualité:

Objectif : Créer des rapports consolidés sur la qualité globale des établissements.

Déroulement Technique du Projet

1. Extraction et Ingestion :

Script Python pour extraire les données de l'API Alim'confiance et les envoyer dans Kafka.

Extraction des avis clients avec une attention à la conformité des données, puis envoi dans Kafka.

2. Traitement avec Spark :

Analyse des sentiments des avis clients via OpenAI.

Enrichissement des données sanitaires et des avis clients avec des informations temporelles et géographiques.

3. Stockage et Indexation avec Elasticsearch :

Stockage des données d'établissement, des niveaux d'hygiène et des avis clients dans Elasticsearch.

4. Visualisation avec Kibana :

Création de tableaux de bord pour suivre la répartition des niveaux d'hygiène et visualiser les sentiments des clients.

Analyses et Indicateurs Attendus

- 1. Indice de Conformité et de Satisfaction : Calculer un score global pour chaque établissement.
- 2. Corrélations entre Hygiène et Sentiment : Identifier si les établissements mal notés pour leur hygiène ont aussi des avis négatifs.
- 3. Zones à Risque : Localiser les zones géographiques où les niveaux d'hygiène et les avis clients sont insatisfaisants.
- 4. Tendances Temporelles : Analyser l'évolution dans le temps des niveaux de satisfaction et d'hygiène.

Exemples de Cas d'Usage - 1. Pour les autorités : Prioriser les contrôles dans les zones ou établissements avec des niveaux d'hygiène et de satisfaction faibles. - 2. Pour les restaurateurs : Identifier les aspects à améliorer pour répondre aux attentes des clients. - 3. Pour les analystes : Suivre les tendances régionales en matière de conformité sanitaire et de satisfaction client.

Déploiement

1. Docker : Conteneurisation des services (Kafka, Spark, Elasticsearch, Kibana) pour simplifier le déploiement et le scaling.
2. Configurations : Variables d'API et paramètres de stockage configurables via des fichiers .env.
3. Automatisation : Un script de déploiement pour exécuter le pipeline complet.

Améliorations Futures

- 1. Machine Learning pour la prédiction des niveaux de conformité.
- 2. Intégration d'autres sources d'avis (réseaux sociaux).
- 3. Développement d'une API pour l'accès aux indicateurs de qualité des établissements.