

Résumé analytique :

Ce projet a été donné dans le cadre d'un test technique de stage. Il vise à répondre aux défis rencontrés par les restaurants dans le traitement efficace des commentaires des clients en mettant en œuvre un système automatisé pour gagner du temps, travailler sur l'engagement des clients et assurer une présence en ligne adéquate.

Les objectifs principaux comprenaient l'analyse du sentiment des commentaires des clients et la génération de réponses appropriées à l'aide de mots-clés de référencement (SEO).

1- Introduction :

Dans le secteur de la restauration, qui évolue rapidement, il est désormais essentiel d'avoir une forte présence digitale pour attirer les clients. Un bon restaurant ne se juge pas seulement sur son service de restauration et sa livraison, mais aussi sur son interactivité en ligne. Les commentaires des clients sont essentiels pour façonner le succès et la réputation des établissements. Cependant, interpréter, comprendre et répondre efficacement aux commentaires est un défi permanent pour les restaurateurs. Ce vise à révolutionner la manière dont les restaurants s'engagent avec leurs clients en ligne en fournissant des réponses significatives aux commentaires sur diverses plateformes.

2- Problématique :

Le secteur de la restauration moderne vise à offrir une expérience complète, de la découverte en ligne à la visite, crucial pour la fidélisation des clients. Gérer les commentaires en ligne est essentiel, mais l'analyse manuelle est chronophage et peut causer des retards ou des erreurs. L'objectif est d'améliorer l'expérience client en évitant ces problèmes.

L'intégration stratégique des mots-clés SEO dans les réponses est souvent négligée, impactant la visibilité en ligne. Un système automatisé, combinant analyse des sentiments, modèle basé sur des règles, et utilisant la NER pour identifier des entités, est nécessaire. Cela permet aux restaurants de comprendre les sentiments des clients et d'améliorer leur présence en ligne de manière efficace.

3- Portée du projet :

Notre projet adopte une approche globale, couvrant divers aspects tels que la collecte des commentaires des clients et l'incorporation stratégique de mots-clés SEO dans la génération de réponses. La portée du projet comprend :

- Curation de mots-clés SEO : La création d'une liste de mots-clés SEO spécifiques à l'industrie afin d'améliorer la visibilité.
- Collecte de données : recueillir les feedbacks des clients pour former la base de données
- Analyse des sentiments : Application de techniques NLP pour classer les commentaires en sentiments.
- Name Entity Recognition (NER) : identifier les entités pertinentes et les sujets sensibles.
- Modèle Rule-based : générer des réponses contextuelles appropriées basées sur le sentiment et l'incorporation stratégique de mots-clés SEO.

4- Méthodologie de travail :

4.1 Collecte des données :

En raison du temps limité qui nous était accordé, nous avons cherché à obtenir un ensemble de données déjà préparé. L'internet est riche en sources ouvertes et fiables d'ensembles de données sur lesquelles nous pourrions potentiellement nous baser. Pour notre problème, nous avons importé de Kaggle un ensemble de données sur les commentaires positifs et négatifs des restaurants.

| Review | Liked |
|-------------------------------------------------------------------|-------|
| Wow... Loved this place. | 1 |
| Crust is not good. | 0 |
| Not tasty and the texture was just nasty. | 0 |
| Stopped by during the late May bank holiday off Rick Steve recomm | 1 |
| The selection on the menu was great and so were the prices. | 1 |
| Now I am getting angry and I want my damn pho. | 0 |
| Honeslty it didn't taste THAT fresh.) | 0 |

Pour Name Entity Recognition, nous avons effectué quelques recherches et créé notre propre ensemble de données contenant des commentaires particuliers utilisant certains mots afin d'aborder les entités dont nous avons besoin.

En utilisant un Annotateur de Reconnaissance d'Entité de Nom open source, nous avons donné les Annotations à nos feedbacks :



C'est ainsi que nous avons généré un fichier json contenant nos données annotées et prêtes à être traitées. Pour ce cas, nous avons choisi de capturer 3 sujets pertinents : Santé, affaire légale et Météo. Cela nous aidera à générer une réponse plus personnalisée

4.2 Analyse des sentiments :

Le modèle d'analyse des sentiments utilise BERT (Bidirectional Encoder Representations from Transformers) par l'intermédiaire de TensorFlow Hub. Le prétraitement des données consiste à nettoyer et à normaliser les critiques, puis à diviser l'ensemble de données. Le modèle traite le texte via le préprocesseur et l'encodeur BERT, remodèle la sortie et la fait passer par une couche LSTM. Une couche d'exclusion empêche l'ajustement excessif, et la prédiction finale utilise une couche dense avec une activation sigmoïde pour la classification binaire des sentiments. Combinant est compilé avec l'optimiseur Adam et la perte d'entropie croisée binaire, offrant une solution robuste pour l'analyse des sentiments en combinant les encastresments BERT et les couches LSTM.

4.3 Named Entity Recognition (NER):

Le modèle NER, mis en œuvre avec spaCy, identifie les entités et les sujets sensibles dans les commentaires des clients. Entraîné sur un modèle spaCy vierge avec des données annotées, un fichier de configuration améliore la précision. Après l'entraînement, le texte d'entrée est traité, générant un dictionnaire d'entités et d'étiquettes identifiées. Cette ressource aide à comprendre les entités et les sujets clés dans les commentaires des clients, offrant une perspective nuancée pour la génération de réponses et l'analyse, aidant ainsi le restaurant à résoudre les problèmes de performance.

4.4 Rule-Based Model:

Le modèle de réponse basé sur des règles élabore des réponses personnalisées en fonction du sentiment, des mots-clés SEO et des mots-clés identifiés. Un sentiment positif déclenche une réponse type avec des mots-clés SEO pertinents. Un sentiment négatif suscite des réponses spécifiques adaptées aux mots-clés tels que la météo, la santé ou les questions juridiques, tout en conservant un ton approprié au contexte. Les commentaires négatifs d'ordre général suscitent des réponses empathiques, exprimant des regrets et un engagement à s'améliorer, avec une personnalisation dynamique à l'aide de mots-clés SEO. Cette approche basée sur des règles garantit des réponses adaptées au sentiment et au contexte pour un engagement significatif des clients.

5- Architecture du système :

L'architecture du système traite efficacement les commentaires des clients, en extrayant les informations clés, en analysant les sentiments et en générant des réponses pertinentes en fonction du contexte. En voici un bref aperçu :

Etape 1 - Input du client et mots-clés SEO : Le processus commence par la saisie des commentaires du client et des mots-clés de référencement pertinents. Cela constitue l'entrée de données initiale pour le système.

Etape 2 - Cartographie des mots-clés : Création d'un dictionnaire à partir des mots-clés SEO saisis, ce qui facilite l'insertion dynamique des mots-clés SEO dans les modèles de réponse.

- Etape 3** - Extraction des entités sensibles : Le système utilise un module NER pour identifier les entités sensibles dans le retour d'information. Cette étape contribue à une analyse plus détaillée et à la personnalisation de la réponse.
- Etape 4** - Analyse des entités du retour d'information : Les entités sensibles identifiées sont ensuite analysées pour déterminer les problèmes ou préoccupations spécifiques. Cette étape est essentielle pour adapter les réponses aux entités identifiées de manière efficace.
- Etape 5** - Préparation de la phrase pour l'analyse de sentiment : La phrase de retour subit un prétraitement pour assurer sa compatibilité avec le modèle d'analyse des sentiments.
- Etape 6** - Analyser le sentiment : Le retour d'information prétraité est introduit dans un modèle d'analyse du sentiment formé, qui prédit le sentiment de l'entrée. Cette étape permet de catégoriser le retour d'information comme positif, négatif.
- Etape 7** - Générer la réponse : Sur la base du sentiment prédit, des entités sensibles identifiées et des mots-clés de référencement, le système utilise un modèle basé sur des règles pour générer une réponse.
- Etape 8** - Sortie : La réponse finale est générée, fournissant une réponse réfléchie et adaptée feedback du client.

6- Résultats et évaluations :

6.1 KPI monitorés :

- Accuracy : le rapport entre le nombre d'instances correctement prédites (vraies positives et vraies négatives) et le nombre total d'instances.
- Precision: le rapport entre les prédictions réellement positives et le nombre total de prédictions positives.
- Recall (or sensitivity): le rapport entre les prédictions réellement positives et le nombre total de prédictions réellement positives. Dans ce cas
- F1-score : la moyenne harmonique de la précision et du rappel. Il fournit un équilibre entre la précision et le rappel. Dans ce cas.
- Matrice de confusion : représentation tabulaire des performances d'un modèle de classification. Elle résume le nombre de prédictions vraies positives, vraies négatives, fausses positives et fausses négatives. Elle est souvent présentée sous la forme d'une matrice, où les lignes correspondent aux classes réelles et les colonnes aux classes prédites.

| Predicted Positive | Predicted Negative |
|--------------------|--------------------|
| Actual Positive | True Positive |
| Actual Negative | False Positive |

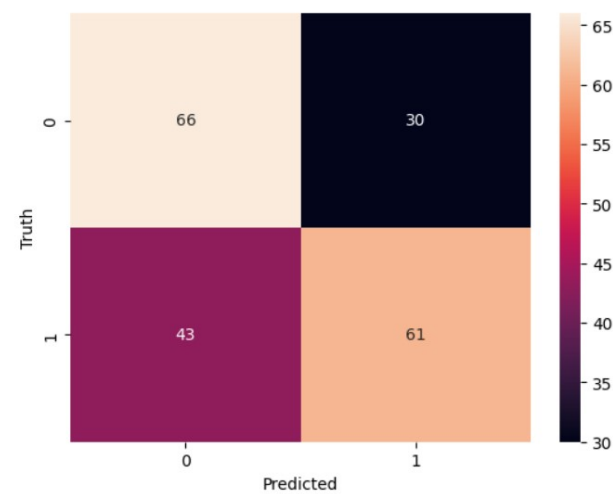
- Loss tok2VEC: utilisée dans le processus de formation d'un modèle basé sur l'algorithme tok2VEC. La fonction de perte quantifie la différence entre les valeurs prédites et les valeurs réelles dans les données d'apprentissage. Il s'agit d'une mesure de la performance du modèle et, pendant la formation, l'objectif est de minimiser cette perte.
- Loss NER: utilisé dans le processus de formation d'un modèle de reconnaissance d'entités nommées (NER). Similaire à la perte tok2VEC, elle mesure la disparité entre les valeurs prédites et les valeurs réelles au cours de la formation. La minimisation de la perte NER est essentielle à la formation d'un modèle NER efficace.

6.2 Sentiment Analysis :

Après quelques recherches et sur la base d'expériences, nous avons choisi de contrôler les performances du modèle par le biais de l'Accuracy , Precision, Recall et du F1-score. le modèle présente des performances raisonnablement équilibrées pour les deux classes, avec une légère préférence pour la classe 0 (classe négative) sur la base des mesures de précision et de rappel.

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.61 | 0.69 | 0.64 | 96 |
| 1 | 0.67 | 0.59 | 0.63 | 104 |
| accuracy | | | 0.64 | 200 |
| macro avg | 0.64 | 0.64 | 0.63 | 200 |
| weighted avg | 0.64 | 0.64 | 0.63 | 200 |

Nous avons régulièrement vérifié la matrice de confusion au cours des expériences pour notre modèle d'analyse. Il s'agit d'un outil puissant qui offre une ventilation détaillée des mesures de performance, révélant les points forts du modèle et les domaines à améliorer. La matrice montre que le modèle est raisonnablement performant, mais qu'il comporte quelques erreurs de classification.



Globalement, les KPIs présentent des résultats intéressants qui sont susceptibles d'être améliorés, mais qui sont également suffisants pour expérimenter.

6.3 Name entity Recognition :

Ce rapport montre la progression des performances du modèle NER au cours de différentes epochs et itérations. Le modèle commence avec des pertes plus élevées mais converge progressivement vers un état où les pertes et les mesures d'évaluation sont optimisées, ce qui indique une reconnaissance efficace des entités nommées.

| E | # | LOSS | TOK2VEC | LOSS | NER |
|-----|-----|------|---------|------|---------|
| 0 | 0 | | 0.00 | | 45.21 |
| 33 | 200 | | 191.76 | | 1222.34 |
| 75 | 400 | | 0.00 | | 0.00 |
| 126 | 600 | | 0.82 | | 0.27 |

6.4 Rule Based Model :

L'évaluation de notre modèle basé sur des règles ne concernait pas le suivi des KPIs, mais plutôt une évolution vers la simulation et l'amélioration manuelle. Il s'agissait d'un effort individuel pour améliorer la personnalisation de la réponse en comparant la façon dont les restaurants ayant une forte visibilité numérique réagissent et en modifiant nos modèles de réponse prédéfinis en conséquence.

7- Outils et technologies utilisés :

NumPy est un Python Package de base pour le calcul scientifique, qui prend en charge les tableaux et les matrices de grande taille, ainsi que les fonctions mathématiques.

Le module Re de Python prend en charge les expressions régulières, souvent utilisées pour le prétraitement de textes, notamment pour supprimer des caractères spéciaux ou des motifs des données d'entrée.

Pour l'analyse des sentiments : TensorFlow et Keras sont essentiels pour construire des réseaux de neurones, TensorFlow Hub accède à des modules pré-entraînés comme BERT, et TensorFlow Text aide au traitement du texte. Pandas s'occupe de la manipulation des données, tandis que Scikit-Learn prend en charge le fractionnement des données pour l'évaluation des modèles. TensorFlow et Keras ont été choisis pour leur facilité d'utilisation, leur flexibilité et l'optimisation des performances dans l'analyse des sentiments. Les couches LSTM sont puissantes pour l'analyse des sentiments, car elles capturent efficacement les dépendances séquentielles.

Pour le NER : spaCy dispose de modèles pré-entraînés pour la reconnaissance d'entités nominales. spaCy Example crée des exemples d'entraînement, et spaCy DocBin stocke efficacement les objets spaCy Doc. tqdm est une bibliothèque Python pour la visualisation rapide des barres de progression dans les tâches, en particulier pour les grands ensembles de données ou les opérations qui prennent beaucoup de temps.

8- Discussion :

Présente le modèle semble répondre à un retour d'information de base en utilisant des mots-clés SEO de base, la solution présente plusieurs limites sur lesquelles nous voudrions certainement travailler.

8-1 Limites du modèle :

- Limites du Rule Based model : En raison de contraintes de temps et de données, notre approche basée sur des règles peut s'avérer difficile à mettre en œuvre dans des situations nuancées, en s'appuyant sur des règles prédéfinies et en nécessitant une intervention manuelle pour l'adaptation.
- Contraintes liées à la compréhension du contexte : La compréhension du modèle est limitée à des règles et des mots-clés prédéfinis, ce qui limite sa capacité à saisir les nuances subtiles ou le sarcasme dans les commentaires des utilisateurs.
- Support multilingue limité : L'approche actuelle n'est peut-être pas optimisée pour les langues autres que l'anglais, ce qui indique qu'il est possible d'améliorer la prise en charge multilingue.
- Gestion des erreurs de grammaire : Le modèle ne gère pas les erreurs de grammaire et de syntaxe, ce qui a un impact sur la Name Entity Recognition.
- Nécessité d'une maintenance permanente : Une maintenance régulière est nécessaire pour que le modèle à base de règles reste aligné sur l'évolution du comportement de l'utilisateur.
- Limites des mots-clés pour le référencement : Bien que les réponses soient orientées vers le référencement, l'utilisation de mots-clés de référencement par le modèle reste limitée et nécessite une intervention humaine.

8-2 Initiatives pour le futur :

- Intégration avec l'API OpenAI : OpenAI fournit des API qui permettent aux développeurs d'interagir avec leurs modèles de langage. Nous pouvons faire des appels API pour générer des réponses basées sur les commentaires de l'utilisateur en utilisant des mots-clés SEO. Les modèles GPT-3 et 4 sont très performants en termes de génération de texte et de contextes conversationnels.
- Support multilingue.
- Recueillir plus de données de feedback
- Amélioration du NER : Nous pouvons affiner la composante NER pour traiter une gamme plus large d'entités et améliorer la précision.
- Analyse avancée de l'engagement de l'utilisateur : Nous pouvons intégrer des analyses avancées pour surveiller l'engagement des utilisateurs dans le système. Il peut s'agir de suivre l'efficacité de la réponse, la satisfaction de l'utilisateur et les interactions avec l'utilisateur afin d'apporter d'autres améliorations.
- Mise à jour dynamique des mots-clés
- Mots clés SEO saisonniers : Nous devons penser à intégrer des mots-clés saisonniers ou temporels qui correspondent à des événements spécifiques, à des vacances ou à des tendances saisonnières. Cela peut améliorer la capacité du système à générer des réponses qui trouvent un écho auprès des utilisateurs pendant des périodes spécifiques.
- Considération des persona d'utilisateurs : Nous devons Tenir compte des persona de votre public cible. (Les préférences linguistiques et les expressions des différents segments d'utilisateurs) afin d'adapter les mots-clés de référencement pour qu'ils trouvent un écho auprès de divers groupes d'utilisateurs.