

Cahier des charges projet MasterCamp :  
**Analyse fine d'avis d'utilisateurs**

FT601- Houssam MOUMOUH

## Table des matières

I.	Contexte de réalisation .....	3
1)	Objectifs de ce projet .....	3
2)	Limites, parties prenantes, exigences particulières .....	3
3)	Besoins et caractéristiques du projet .....	4
II.	Problématique et objet du projet .....	5
1)	Données existantes, formes et description .....	5
2)	Contraintes humaines et matérielles .....	6
3)	Objectifs premiers et finaux du projet .....	6
4)	Répartition des tâches au sein de l'équipe .....	6
5)	Description des étapes d'exécution .....	7
III.	Bilan et perspectives .....	8
1)	Le rendu attendu .....	8
2)	Fonctionnalités et ouverture pour d'éventuelles reprises ou améliorations.....	8
3)	Documentation du projet.....	9
IV.	Conclusion .....	10
V.	Annexes .....	11
1)	UML diagramme .....	12

# I. Contexte de réalisation

## 1) Objectifs de ce projet

L'objectif de ce projet est d'aider les entreprises à s'améliorer au niveau de leurs produits et de leur service à travers une analyse des avis de leurs clients. En effet, cela peut être long et fastidieux de voir les avis un à un afin d'en relever les informations pertinentes. Notre objectif est donc de les aider à les relever de façon automatique. L'analyse des avis doit permettre aux entreprises de comprendre le ressenti des utilisateurs à travers leurs sentiments exprimés dans les commentaires.

Ainsi, les entreprises peuvent bénéficier d'une meilleure compréhension des réactions des consommateurs, ce qui leur permet d'améliorer leurs produits ou service.

## 2) Limites, parties prenantes, exigences particulières

### **Limites :**

Il existe différentes limites à notre projet.

Dans un premier temps, nous dépendons des avis mis à notre disposition. Notre système est donc limité par notre base de données. Ainsi, la représentativité des résultats dépend donc du contenu de celle-ci.

Ensuite, la précision de l'analyse des sentiments est complexe et limitée. En effet, une émotion peut être compliquée à reconnaître pour une intelligence artificielle qui aura du mal à détecter le sarcasme ou l'ironie, et à décrypter les messages ambigus.

Une autre limite de notre projet est la difficulté de travailler avec plusieurs langues. Si un système est conçu pour fonctionner avec plusieurs langues, assurer la même précision dans toutes les langues peut devenir plus complexe car il faut s'adapter à chaque langue et chaque culture. Les idiosyncrasies linguistiques et culturelles peuvent affecter la capacité d'un système à analyser correctement les commentaires dans différentes langues.

Une dernière limite est le fait que l'on se base sur un seul type d'avis. En effet, nous nous limitons aux avis comportant des étoiles et un commentaire.

### **Parties prenantes :**

Les parties prenantes externes sont les utilisateurs de notre système qui sont des entreprises, et les clients de ces entreprises. En effet, les entreprises, qui sont des acteurs actifs, utilisent notre système afin d'améliorer leurs produits ou service. Ensuite, les clients de ces entreprises, qui sont des acteurs passifs, laissent des avis sur les produits ou services utilisés qui sont analysés par notre système. De plus, notre système a des retombées sur les clients qui pourront être d'autant plus satisfaits si les entreprises s'améliorent en fonction de leurs avis. Aussi, l'EFREI est une partie prenante externe car l'école nous accompagne dans la réalisation de ce projet.

Les parties prenantes internes sont Lionel PIMENTA SILVA qui est le chef de projet ainsi que Marion NGUYEN, Edwin SAVORY et Morgan SENECHAL qui collaborent sur la réalisation du projet.

#### **Exigences particulières :**

En termes de performance, le système doit être capable de gérer efficacement de larges volumes d'avis d'utilisateurs, permettant ainsi une analyse rapide et précise. Il doit également être adaptatif pour pouvoir s'adapter à une augmentation du nombre d'avis au fil du temps.

L'interface de l'utilisateur revêt une importance capitale. Elle doit être intuitive pour permettre aux utilisateurs de naviguer facilement et de lire facilement les résultats. Aussi cela est essentiel pour permettre aux utilisateurs de tirer pleinement parti du système.

### **3) Besoins et caractéristiques du projet**

#### **Besoins du projet :**

Ce projet a plusieurs étapes importantes tels que la collecte des avis des utilisateurs, le prétraitement des données, l'analyse des sentiments, la catégorisation des avis et l'interface utilisateur.

En effet, pour recueillir les avis des utilisateurs sur le produit ou le service en question, nous devons effectuer une collecte de ces avis à l'aide de techniques telles que le web scraping (extraction du contenu d'un site web), les API ou en se connectant à une plateforme d'avis en ligne.

Une fois ces avis recueillis, nous devons entamer un prétraitement des données afin de supprimer les informations superflues telles que les balises HTML, la ponctuation et les mots vides. De plus, il faudra standardiser les mots pour homogénéiser les données.

Après avoir prétraité nos données, notre système doit être capable d'analyser les sentiments exprimés dans les avis. Cela nécessitera donc l'utilisation de techniques d'analyse des sentiments telles que les dictionnaires de sentiments, les modèles d'apprentissage automatique et les réseaux de neurones.

Ensuite, il faudra catégoriser nos avis selon les catégories prédéfinies (positif, négatif, neutre) ou selon des aspects spécifiques des produits et services (qualité, prix, service client, etc.). Cela nous permettra de mieux analyser les avis des consommateurs et de mieux comprendre les opportunités d'amélioration.

Enfin, une interface utilisateur intuitive est essentielle pour que les utilisateurs (entreprises) puissent utiliser le système facilement. L'interface doit pouvoir afficher les résultats de l'analyse, filtrer les avis par catégorie ou aspect, et fournir des outils d'exportation et de création de rapports pour faciliter l'utilisation des informations extraites.

Ainsi, en répondant à ces besoins, le projet fournit une solution complète et efficace pour analyser les avis des utilisateurs, permettant aux entreprises d'améliorer leurs produits et services en fonction des réactions et des sentiments des consommateurs.

#### **Caractéristiques du projet :**

Lors du développement d'un système d'analyse des sentiments et de classification des commentaires, plusieurs aspects clés doivent être pris en compte pour garantir son efficacité et sa valeur pour les entreprises. La précision et fiabilité, adaptatif, la flexibilité, le temps de réponse rapide, la sécurité des données et la documentation et support sont des éléments clés garantissant un système d'analyse performant et fiable, répondant aux besoins des entreprises.

En effet, la précision et fiabilité est un aspect important de notre système qui doit être précis dans l'analyse des sentiments exprimés et dans la classification des commentaires. Une attention particulière doit être portée à la qualité des résultats afin de fournir des informations fiables et pertinentes pour les entreprises.

D'une part, notre système doit être adaptatif pour traiter un volume croissant de commentaires des utilisateurs au fil du temps. Il doit être capable de gérer efficacement de grandes quantités de données sans compromettre les performances.

D'autre part il doit être flexible pour s'adapter à différentes langues. Il doit permettre la personnalisation des lexiques de sentiments, des catégories de classification et des aspects à analyser.

De plus, la performance concernant le temps de réponse doit être optimisé. Le système doit fournir des résultats d'analyse rapidement, en minimisant le temps de traitement des avis des utilisateurs. Les entreprises doivent pouvoir obtenir des informations rapidement pour prendre des décisions informées.

La sécurité des données des utilisateurs reste aussi un aspect primordial à ce genre de système. Le projet doit mettre en place des mesures de sécurité pour protéger les informations confidentielles et garantir la conformité aux réglementations en matière de protection des données.

Enfin, une documentation et un support doit être mis à disposition de nos utilisateurs. Cela facilitera l'utilisation et la compréhension du système, ainsi que la résolution des problèmes potentiels.

## II. Problématique et objet du projet

### 1) Données existantes, formes et description

Les données existantes pour ce projet sont les commentaires des utilisateurs sur les produits. Ces commentaires peuvent être trouvés sur diverses plateformes en ligne telles que les sites d'e-commerce, les réseaux sociaux, les forums, les blogs, etc. Les commentaires peuvent être sous forme de texte libre où les utilisateurs expriment leurs opinions, leurs expériences et leurs sentiments concernant les produits.

Les données seront collectées à partir de sources en ligne en utilisant des techniques de scraping web ou des API disponibles pour extraire les commentaires des utilisateurs. Les commentaires seront généralement structurés en texte brut avec des informations supplémentaires telles que l'identifiant du produit, la date du commentaire, etc.

Notes attribuées : Certains commentaires peuvent être accompagnés de notes ou d'évaluations numériques attribuées par les utilisateurs pour exprimer leur satisfaction ou leur mécontentement.

Ces notes peuvent être sur une échelle numérique, par exemple de 1 à 5 étoiles, ou toute autre échelle définie par la plateforme.

## 2) Contraintes humaines et matérielles

Les principales contraintes humaines pour ce projet sont d'avoir une équipe de data scientists compétents en programmation Python et en analyse de données. Ils devront maîtriser les techniques de prétraitement de texte, de traitement du langage naturel et d'apprentissage automatique pour développer le système d'analyse de sentiment.

Du côté matériel, il faudra disposer d'ordinateurs avec suffisamment de puissance de calcul pour traiter les grandes quantités de données et effectuer des analyses en temps opportun.

## 3) Objectifs premiers et finaux du projet

Les objectifs premiers du projet consistent à développer un système d'analyse de commentaires utilisateurs en utilisant des techniques de traitement du langage naturel et d'apprentissage automatique. Ce système devra être capable d'extraire les sentiments exprimés dans les commentaires et de les classer en sentiments positifs, négatifs ou neutres.

L'objectif final est d'aider les entreprises à mieux comprendre les réactions des consommateurs envers leurs produits ou services. En analysant les sentiments exprimés, les entreprises pourront obtenir des informations précieuses sur les aspects spécifiques de leurs produits qui suscitent des réactions positives ou négatives. Cela leur permettra d'améliorer leurs produits ou services en fonction des retours des utilisateurs.

## 4) Répartition des tâches au sein de l'équipe

La répartition des tâches au sein de l'équipe peut se faire de la manière suivante :

- **Collecte des données** : Un membre de l'équipe sera responsable de la collecte des commentaires des utilisateurs à partir des sources en ligne en utilisant des techniques de scraping web ou des API.
- **Prétraitement des données** : Un autre membre de l'équipe sera chargé de nettoyer et prétraiter les données en supprimant les caractères indésirables, les fautes d'orthographe, les doublons, etc. Il effectuera également une normalisation du texte en convertissant en minuscules et en supprimant les mots vides (stop words).
- **Développement de modèles** : Un data scientist spécialisé en traitement du langage naturel et en apprentissage automatique sera responsable du développement de modèles d'analyse de sentiment en utilisant des techniques telles que les modèles de sacs de mots (bag-of-words), les modèles basés sur les réseaux de neurones, ou les modèles transformer comme BERT.

- **Mise en œuvre du système** : Un autre membre de l'équipe avec des compétences en développement Python sera chargé de mettre en œuvre le système d'analyse de commentaires utilisateurs en utilisant les modèles développés. Il pourra utiliser des bibliothèques Python telles que NLTK, scikit-learn et TensorFlow pour l'implémentation.
- **Responsable du projet** : Supervise l'ensemble du projet, coordonne les différentes tâches et assure le respect des délais.

## 5) Description des étapes d'exécution

Les étapes d'exécution du projet peuvent être les suivantes :

- **Collecte des données** : Utilisation de techniques de scraping web ou d'API pour extraire les commentaires des utilisateurs à partir des sources en ligne.
- **Prétraitement des données** : Nettoyage des données en supprimant les caractères indésirables, les fautes d'orthographe et en normalisant le texte.
- **Exploration des données** : Analyse exploratoire des données pour comprendre leur distribution, identifier les tendances et les schémas.
- **Préparation des données** : Transformation des données textuelles en représentations numériques adaptées à l'analyse, telles que les vecteurs de mots ou les représentations basées sur les réseaux de neurones.
- **Développement de modèles** : Utilisation de techniques d'apprentissage automatique pour développer des modèles d'analyse de sentiment en utilisant les données prétraitées.
- **Évaluation des modèles** : Évaluation de la performance des modèles en utilisant des mesures telles que la précision, le rappel et la F-mesure. Sélection du modèle le plus performant.
- **Mise en œuvre du système** : Intégration du modèle sélectionné dans un système Python qui peut traiter les commentaires des utilisateurs et prédire les sentiments.
- **Validation et ajustement** : Validation du système en utilisant des données supplémentaires ou des données réelles pour évaluer sa performance en conditions réelles. Effectuer des ajustements si nécessaire pour améliorer la précision et l'efficacité du système.
- **Rapport final et recommandations** : Présentation des résultats de l'analyse de sentiment, fourniture d'informations et de recommandations aux entreprises pour améliorer leurs produits ou services en fonction des réactions des utilisateurs.

Il est important de noter que ces étapes peuvent être itératives et nécessiter des ajustements en fonction des résultats obtenus au cours du processus.

### III. Bilan et perspectives

#### 1) Le rendu attendu

Le rendu attendu peut inclure les éléments suivants :

- **Application fonctionnelle** : Une application ou une interface utilisateur conviviale permettant aux utilisateurs (les entreprises) d'interagir avec le système. L'application devrait offrir des fonctionnalités telles que l'importation des commentaires des utilisateurs, l'analyse des sentiments, la classification des commentaires, la visualisation des résultats et la génération de rapports.
- **Analyse des sentiments précise** : Le système devrait être capable de détecter et d'analyser avec précision les sentiments exprimés dans les commentaires des utilisateurs. Il devrait être capable de classer les commentaires en fonction des sentiments (positif, négatif, neutre) et de fournir une évaluation globale des réactions des consommateurs.
- **Classification des commentaires** : Le système devrait être en mesure de classer les commentaires dans des catégories spécifiques, permettant aux entreprises de mieux comprendre les aspects spécifiques de leurs produits ou services qui suscitent des réactions positives ou négatives chez les consommateurs. Par exemple, la classification peut inclure des catégories telles que la qualité, le prix, la livraison, le service client, etc.
- **Visualisation des résultats** : L'application devrait fournir des moyens de visualisation des résultats d'analyse, tels que des graphiques, des tableaux ou des représentations visuelles, pour faciliter la compréhension et l'exploration des réactions des consommateurs. Les utilisateurs devraient être en mesure de filtrer les commentaires par catégorie, de rechercher des mots clés spécifiques et de générer des rapports personnalisés.
- **Documentation et guide d'utilisation** : Un document de documentation complet devrait accompagner l'application, fournissant des instructions détaillées sur l'installation, la configuration et l'utilisation du système. Un guide d'utilisation clair et concis devrait être fourni pour aider les utilisateurs à tirer le meilleur parti de l'application.

#### 2) Fonctionnalités et ouverture pour d'éventuelles reprises ou améliorations

Les fonctionnalités du projet peuvent être définies en fonction des objectifs et des besoins identifiés. Voici quelques fonctionnalités potentielles pour un système d'analyse des commentaires des utilisateurs :

- **Collecte des données** : Permettre la collecte automatisée des commentaires à partir de différentes sources, telles que des plateformes d'avis en ligne, des réseaux sociaux, des forums, etc. Cette fonctionnalité peut inclure la récupération et le stockage des données dans une base de données.
- **Prétraitement des données** : Effectuer un prétraitement des données pour nettoyer le texte, supprimer les caractères indésirables, corriger les fautes d'orthographe, normaliser le texte, etc. Cela garantit des données cohérentes et de meilleure qualité pour l'analyse.



- **Analyse des sentiments** : Utiliser des techniques de traitement du langage naturel et d'apprentissage automatique pour analyser les sentiments exprimés dans les commentaires. Cela peut inclure la détection des opinions positives, négatives ou neutres, l'identification des aspects spécifiques mentionnés, et la classification des commentaires dans des catégories pertinentes.
- **Visualisation des données** : Fournir des visualisations graphiques et des rapports pour représenter les résultats de l'analyse des commentaires. Cela peut inclure des graphiques à barres, des nuages de mots, des cartes de chaleur, des tableaux de bord interactifs, etc., pour permettre une compréhension facile des tendances et des insights.
- **Gestion des utilisateurs et des produits** : Permettre aux entreprises de gérer les utilisateurs, les produits ou les services spécifiques faisant l'objet des commentaires. Cela peut inclure la possibilité de filtrer et d'organiser les commentaires par produit, d'identifier les utilisateurs les plus actifs, etc.
- **Notifications et alertes** : Envoyer des notifications ou des alertes aux entreprises lorsqu'un commentaire critique ou problématique est détecté, leur permettant de prendre des mesures immédiates pour résoudre les problèmes.
- **Personnalisation et adaptabilité** : Offrir des options de personnalisation pour ajuster les paramètres d'analyse des sentiments en fonction des besoins spécifiques de l'entreprise. Cela peut inclure la possibilité de définir des listes de mots clés, des dictionnaires de sentiments personnalisés, des filtres de langue, etc.

Pour l'ouverture aux reprises ou améliorations futures, certaines fonctionnalités supplémentaires peuvent être envisagées :

- **Analyse de la tonalité** : Développer des fonctionnalités pour détecter la tonalité générale des commentaires, telle que l'ironie, le sarcasme, l'enthousiasme, etc.
- **Analyse des entités et des aspects** : Identifier et extraire les entités (produits, lieux, personnes, etc.) mentionnées dans les commentaires, ainsi que les aspects spécifiques associés à ces entités.
- **Analyse comparative** : Permettre la comparaison des commentaires entre plusieurs produits ou services similaires pour fournir des informations sur la perception comparative des consommateurs.
- **Analyse des tendances** : Identifier les tendances émergentes et les changements de sentiments au fil du temps pour permettre aux

### 3) Documentation du projet

Pour faciliter la compréhension des différentes fonctionnalités du système par les utilisateurs, nous prévoyons d'implémenter un fichier Readme.txt qui permettra aux utilisateurs de trouver rapidement des informations sur l'utilisation de notre système.

## IV. Conclusion

**Opinion fi** est donc l'application idéale pour les entreprises qui ne veulent pas perdre de temps à épier tous les avis avant de devoir prendre une décision cruciale quant à l'amélioration de leurs produits ou de leurs services. **Opinion fi** permet une économie de temps incomparable et grâce à sa facilité d'utilisation. De plus, tout est totalement gratuit.

## V. Annexes

### 1) Use case Diagram

Ici nous pouvons voir le diagramme impliquant les différents acteurs ayant un lien direct avec le fonctionnement du produit. On y trouve évidemment le client, l'application mais également PayPal. Notre business model fonctionnant à travers les dons, PayPal est un acteur majeur de notre projet.

