



Evaluation Certifiante



Identification des besoins pour la mise en place d'une solution d'analyse automatisée des tweets clients adressés à une entreprise de télécommunications

Bloc 2 - C2.1



Réalisé par :

SGHIOURI Mohammed

SOBGUI Ivan Joel

BOTI Armel Cyrille

BEN LOL Oumar

DIVENGI KIBANGUDI BUNKEMBO Nagui

ELOUMOU MBOUDOU Pascal Aurele

Date : 19 Novembre 2025

Tables de matières

Introduction.....	2
1. Présentation du commanditaire.....	2
1.1. Profil fictif du client.....	2
1.2. Organisation actuelle du service après-vente.....	2
2. Identification des besoins.....	2
2.1. Besoins d'automatisation.....	2
2.2. Pertinence d'une solution NLP + LLM.....	3
2.3. Attentes vis-à-vis du traitement des tweets.....	3
3. Analyse initiale du fichier free tweet export.csv.....	3
3.1. Description des variables.....	3
3.2. Nature des données textuelles.....	3
3.3. Techniques de nettoyage et preprocessing NLP.....	4
3.4. Identification de défis techniques.....	4
4. Définition de personas clients et utilisateurs internes.....	5
4.1. Personas clients.....	5
4.2. Utilisateurs internes.....	5
5. Synthèse de la problématique.....	6
5.1. Formulation de la problématique SAV.....	6
5.2. Objectif général du projet.....	6
5.3. Périmètre et limites.....	6
6. Conclusion.....	6
7. Références et sources.....	8

Introduction

Les entreprises de télécommunications font face à un volume croissant de messages clients sur les réseaux sociaux, nécessitant des solutions de traitement automatisé pour maintenir la qualité du service. Dans ce contexte, cette étude vise à identifier et structurer les besoins de Free en matière d'analyse automatisée de tweets clients via LLM, afin de traiter, classifier et analyser les messages en vue d'extraire des indicateurs clés.

1. Présentation du commanditaire

1.1. Profil fictif du client

Free, filiale du groupe Iliad, représente l'un des principaux opérateurs de télécommunications en France. L'entreprise reçoit quotidiennement plusieurs milliers de tweets mentionnant ses services, son service client ou ses produits. Les messages sont souvent du type plaintes, demandes d'information, commentaires et réclamations. Ce qui constitue une source d'information précieuse, à condition qu'ils puissent être traités de manière efficace et structurée.

1.2. Organisation actuelle du service après-vente

Structure traditionnelle avec équipe d'agents. Service opérationnel généralement de 8h à 20h en semaine. Standards de réponse souvent inférieurs à 1h pour les demandes urgentes. L'analyse du fichier free tweet export.csv révèle des frustrations face aux délais : *"13 minutes 13 sec!!!! c'est le temps ENCOREEEEE UNE FOIS PERDU à attendre après free pour résoudre un problème récurrent !! INCOMPÉTENCE !!!!"* (tweet ID: 1813254932449140912).

Le traitement manuel présente des limitations en scalabilité. Les clients utilisent Twitter comme canal de réclamation public, amplifiant l'impact négatif des retards.

2. Identification des besoins

2.1. Besoins d'automatisation

L'analyse du fichier free tweet export.csv (5 000 tweets) révèle une diversité de messages : plaintes techniques urgentes, problèmes de connexion critiques, demandes d'information, retweets promotionnels, messages de satisfaction.

Tri et classification : distinguer automatiquement les catégories pour optimiser le routage et prioriser les demandes urgentes.

Détection de sujets : identifier automatiquement les thématiques (problèmes de connexion, forfaits mobiles, équipements Freebox, résiliation, facturation) pour améliorer le routage et générer des insights.

Analyse de sentiment : déterminer le ton émotionnel (positif, négatif, neutre) et identifier les situations d'urgence. Les tweets présentent des expressions variées, avec emojis négatifs (😞, 😡, 🤬) et langage émotionnel fort.

Détection d'urgences : identifier automatiquement les situations nécessitant une intervention immédiate (coupures de service majeures, problèmes de sécurité, réclamations virulentes).

2.2. Pertinence d'une solution NLP + LLM

Les modèles modernes (Gemini, GPT, Mistral, Claude) comprennent le langage naturel informel, adapté aux tweets, on peut citer le langage familier, les abréviations ou encore les emojis. Aussi, ils ont une approche flexible qui s'adapte aux évolutions du langage et qui capture les nuances subtiles et les contextes complexes. Ils offrent aussi la possibilité de fine-tuning sur données spécifiques avec extraction automatique d'informations clés et génération de résumés structurés.

2.3. Attentes vis-à-vis du traitement des tweets

Aide à la décision : fournir aux décideurs des informations structurées et actionnables (détection de recrudescence de plaintes, alertes pour intervention proactive).

Visualisation : transformer les données textuelles en représentations graphiques interactives (tableaux de bord avec métriques clés : volume par catégorie, évolution du sentiment, répartition par thématique).

Pré-classement : catégoriser automatiquement chaque tweet (type de demande, urgence, sentiment, thématique) pour faciliter le travail des agents et optimiser le routage.

Le traitement automatisé ajoute trois colonnes au dataset : sentiment (negative/neutral/positive), priority (0/1/2), main_topic (sujet principal identifié).

3. Analyse initiale du fichier free tweet export.csv

3.1. Description des variables

Variables principales : id (identifiant unique), created_at (horodatage), full_text (contenu textuel), variables utilisateur (screen_name, name, user_id), métadonnées (in_reply_to, retweeted_status, quoted_status), engagement (favorite_count, retweet_count, views_count), média (media, media_tags).

3.2. Nature des données textuelles

Hétérogénéité linguistique : mélange de langage formel et informel, français standard et expressions familières, abréviations (“jpp”, “ftth”), parfois anglais. Exemples : langage technique (“Erreur 20 d’authentification”, “FTTH”), langage familier (“jpp”, “ça déconne”).

Présence de bruit : caractères spéciaux, emojis, formats non standardisés. Proportion significative de retweets promotionnels ne nécessitant pas de traitement par le service client.

Emojis et symboles : fréquents, véhiculant des informations sur le sentiment. Techniques de preprocessing : normalisation des emojis (😞 → “visage triste”), gestion des caractères spéciaux, préservation du contexte émotionnel.

Mentions et hashtags : mentions multiples fréquentes (@free, @Freebox, @Free_1337), hashtags (#Free, #fibre, #FTTH) fournissant des informations contextuelles.

3.3. Techniques de nettoyage et preprocessing NLP

Normalisation du texte : suppression des URLs (t.co), gestion des mentions (@username), préservation des hashtags (#Free, #fibre), normalisation des répétitions (“ENCOREEEEE” → “ENCORE”).

Gestion des emojis : conversion en descriptions textuelles pour préserver l’information sentimentale, nettoyage des caractères spéciaux, normalisation des espaces.

Traitement du langage informel : préservation des abréviations techniques (FTTH, IPv6, Freebox), normalisation optionnelle des abréviations courantes (“jpp” → “j’en peux plus”), gestion des fautes de frappe.

3.4. Identification de défis techniques

Langage familier et expressions idiomatiques : registre très informel, expressions spécifiques à la culture française et aux réseaux sociaux.

Ironie et sarcasme : fréquents dans les plaintes, nécessitant une détection correcte pour éviter des erreurs d’interprétation.

Demandes multiples par tweet : un tweet peut contenir plusieurs demandes distinctes, compliquant la classification et le routage.

Contexte implicite : tweets courts pouvant faire référence à des situations non explicitement mentionnées.

Variabilité orthographique et grammaticale : fautes de frappe, abréviations, conventions Twitter (majuscules pour emphase, répétition de caractères).

Volume et scalabilité : 5 000 tweets analysés, mais en production, Free pourrait recevoir des dizaines de milliers de tweets par jour, nécessitant un traitement en temps réel.

4. Définition de personas clients et utilisateurs internes

4.1. Personas clients

Client tech-savvy : jeune, familier avec les technologies, tweets précis et techniques (FTTH, IPv6), attend des réponses rapides et techniques.

Client en colère : utilise Twitter comme canal de réclamation public, langage émotionnel fort, emojis négatifs, menaces de résiliation. Nécessite une attention particulière et une réponse rapide.

Client demandeur d'information : pose des questions sur services, tarifs, fonctionnalités, procédures. Tweets souvent formulés sous forme de questions directes.

Client satisfait : exprime son contentement ou partage des expériences positives. Source précieuse d'informations sur les points forts.

Client en situation d'urgence : signale des problèmes critiques nécessitant une intervention immédiate (coupure de service, problème de sécurité). Nécessite une détection prioritaire et un routage automatique.

4.2. Utilisateurs internes

Agents du service client : utilisateurs primaires, utilisent pré-classement et routage pour identifier rapidement les tweets nécessitant intervention. Interface intuitive, rapide, informations actionnables.

Managers du service client : utilisent la solution pour pilotage et visualisation. Besoin de tableaux de bord avec métriques clés (volume par catégorie, temps de réponse moyen, évolution du sentiment).

Analystes data : analyses approfondies, identification de patterns, analyses de tendances, rapports détaillés. Besoin d'export de données, API pour intégration, fonctionnalités d'analyse avancées.

Décideurs stratégiques : aide à la décision, identification de problèmes récurrents, évaluation d'impact, anticipation de crises. Besoin de synthèses claires, alertes sur situations critiques, insights actionnables.

Gestion de l'authentification : système d'authentification JWT, gestion des utilisateurs avec contrôle d'accès basé sur les rôles (administrateur, analyste de données), permissions granulaires.

5. Synthèse de la problématique

5.1. Formulation de la problématique SAV

Comment automatiser efficacement le traitement, la classification et l'analyse des tweets clients adressés à Free sur Twitter, en exploitant les capacités des modèles de langage modernes, afin d'améliorer la réactivité du service client, d'optimiser le routage des demandes, et de générer des insights actionnables pour la prise de décision stratégique, tout en faisant face aux défis spécifiques du langage informel, de l'ironie, et de la multiplicité des demandes par message ?

5.2. Objectif général du projet

Poser les bases méthodologiques et techniques d'une solution automatisée de traitement des tweets clients, en vue de son prototypage et de son déploiement ultérieur. Objectifs spécifiques : structuration des besoins, définition de l'architecture technique, identification des contraintes et risques, définition des métriques de succès.

5.3. Périmètre et limites

Le présent travail se concentre exclusivement sur la phase d'identification des besoins et de structuration de la problématique. Les phases ultérieures telles que le prototypage, le développement, le déploiement, l'analyse des résultats et recommandations opérationnelles feront l'objet des travaux des blocs CC2.2 et CC2.3.

6. Conclusion

Cette étude a structuré une problématique multidimensionnelle, mettant en évidence les enjeux techniques, opérationnels et stratégiques.

Les besoins de Free sont multiples : tri/classification automatique, détection de thématiques, analyse de sentiment, détection d'urgences. Ces besoins justifient l'utilisation d'une solution NLP + LLM pour faire face aux défis du langage informel, de l'ironie et de la multiplicité des demandes.

L'examen du fichier free tweet export.csv (5 000 tweets) révèle : hétérogénéité linguistique, bruit, emojis, mentions/hashtags. Diversité de messages : plaintes techniques, problèmes critiques, demandes d'information, retweets, messages de satisfaction.

Les défis techniques (langage familier, ironie, demandes multiples, contexte implicite, variabilité orthographique, volume) définissent un contexte complexe nécessitant une approche rigoureuse. Les techniques de nettoyage NLP + capacités LLM permettront de traiter cette diversité.

La définition des personas clients et utilisateurs internes a précisé les besoins spécifiques. La synthèse de la problématique formule clairement l'objectif : poser les bases méthodologiques et techniques pour prototypage et déploiement.

7. Références et sources

- Fichier free tweet export.csv : 5 000 tweets collectés mentionnant Free (période 2020-2024)
- Large Language Models in Telecommunications :
<https://arxiv.org/html/2308.06013v2>
- BERTweetFR : A French Tweet-Specific Language Model :
<https://arxiv.org/abs/2109.10234>