



Evaluation Certifiante

Bloc CC2.1

Transformer le flux Twitter en actions : Une solution IA pour optimiser le SAV Free Mobile

Réalisé par :

SGHIOURI Mohammed

SOBGUI Ivan Joel

BOTI Armel Cyrille

BEN LOL Oumar

DIVENGI KIBANGUDI BUNKEMBO

Nagui

ELOUMOU MBOUDOU Pascale Aurele

Github repository

Date : 19 Novembre 2025

Table des matières

1	Architecture Technique et Pipeline de Traitement	3
1.1	Structure du Projet et Services Clés	3
1.2	Architecture de la Base de Données (MongoDB)	3
1.3	Pipeline de Nettoyage et Structuration du Dataset	4
1.4	Supervision et Contrôle d'Administration	4
2	Définition et Calcul des KPI SAV	7
2.1	KPI de Volume et de Flux	7
2.2	KPI de Satisfaction et d'Urgence	7
2.3	KPI d'Engagement et Viralité	7
3	Exploration Visuelle et Interprétation	8
3.1	Analyse Temporelle et Sentimentale	8
3.2	Analyse Croisée (Topics vs Sentiments)	9
4	Qualité des Données et Limites du Modèle Actuel	10
5	Conclusion Technique	11

Table des figures

1	Architecture Initial	4
2	Gestion des rôles	5
3	Interface de gestion des Utilisateurs, illustrant le contrôle d'accès basé sur les permissions.	5
4	Interface de gestion des Rôles, illustrant le contrôle d'accès basé sur les permissions.	5
5	Vue des Logs Système et Événements, fournissant une traçabilité détaillée des actions et des statuts de tâche.	6
6	La page d'accueil de notre application	6
7	La page d'accueil de notre application	6
8	Interface de Configuration des Modèles IA, illustrant la modularité pour le choix de l'API ou du modèle local.	7
9	Tableau de Bord Analytique Global, présentant une vue synthétique des KPI de performance du système et de satisfaction client.	8
10	Timeline de l'Évolution des Sentiments, outil clé pour l'identification des pics d'insatisfaction et leur corrélation temporelle.	8
11	Timeline de l'Évolution des Sentiments, outil clé pour l'identification des pics d'insatisfaction et leur corrélation temporelle.	9
12	Distribution croisée des Topics et Priorités, mettant en lumière les thèmes critiques nécessitant une intervention immédiate.	9

Liste des tableaux

1	Composants principaux de l'architecture Dallosh Analysis.	3
2	Rôle des collections clés dans MongoDB.	3

1 Architecture Technique et Pipeline de Traitement

Notre solution, *Dallosh Analysis*, repose sur une architecture **microservices** conçue pour garantir la flexibilité, l'évolutivité et une supervision en temps réel.

1.1 Structure du Projet et Services Clés

Le système est orchestré par `Docker/Docker Compose` et se divise en trois composants principaux :

Composant	Rôle Fonctionnel	Technologies Associées
frontend	Interface Utilisateur, Visualisation des KPI	Next.js, React, Ampq
backend	API REST (Authentification, Fichiers, Tâches), DB	TypeScript, Node.js, MongoDB
microservices	Cœur du traitement ETL, Asynchrone	Python, Celery, RabbitMQ

TABLE 1 – Composants principaux de l'architecture Dallosh Analysis.

Rôle des Microservices (Python) :

- `reading_file.py` : Gère l'ingestion du fichier CSV brut en `DataFrame Pandas`.
- `cleaning.py` : Applique les règles de filtrage (`Regex`, `IQR`) pour la déduplication et l'assainissement du texte (voir Section 1.3).
- `calling_llm.py` : Orchestre l'appel au LLM pour l'enrichissement (*Sentiment*, *Topic*, *Priorité*), incluant la gestion des tentatives (*retry*) et de la pagination.
- `processor.py` : Module Celery garantissant l'ordonnancement des étapes et le suivi des statuts (`TASK_STATUS_DONE`, etc.) via le *message broker RabbitMQ*.

1.2 Architecture de la Base de Données (MongoDB)

Le choix d'une base de données **NoSQL** (MongoDB) permet une gestion agile des schémas de données pour les métadonnées et la configuration.

Collection	Description	Fonction Critique
<code>files</code>	Métadonnées des datasets uploadés et de leurs chemins.	Traçabilité des tâches et des sources.
<code>tasks</code>	Historique et statut des pipelines exécutés.	Supervision temps réel de l'avancement (<code>RabbitMQ</code>).
<code>settings</code>	Configuration modulaire des modèles LLM (clés, limites, mode local/externe).	Adaptabilité aux contraintes externes et aux coûts API.
<code>users</code> et <code>roles</code>	Informations d'authentification et permissions.	Sécurité et contrôle d'accès aux fonctionnalités.

TABLE 2 – Rôle des collections clés dans MongoDB.

1.3 Pipeline de Nettoyage et Structuration du Dataset

Le traitement des données suit une logique de pré-traitement stricte avant l'enrichissement par l'Intelligence Artificielle.

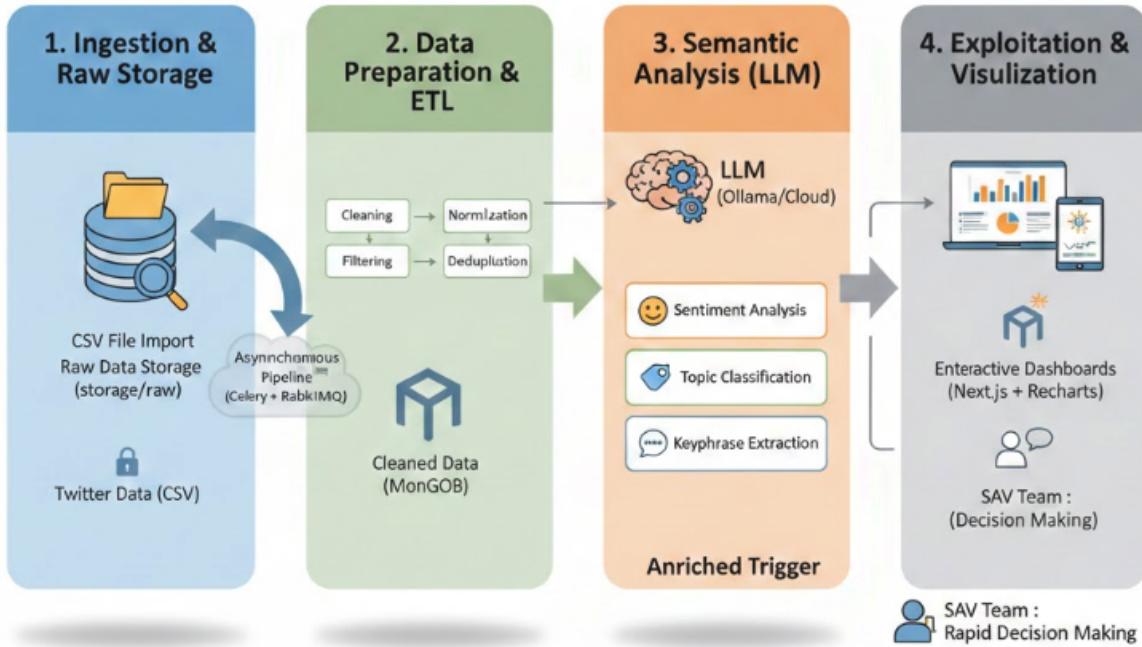


FIGURE 1 – Architecture Initial

Ingestion & Nettoyage Avancé :

- Les émojis sont supprimés via Regex Unicode pour garantir la pureté de l'analyse sémantique.
- Les mentions (@user) sont préservées par masquage temporaire (_MENTION_i_) pour éviter la corruption du texte lors du nettoyage des caractères spéciaux.
- La Déduplication et la suppression des valeurs aberrantes (outliers) sont assurées par une méthode statistique robuste (Interquartile Range - IQR).

Enrichissement par IA (Mode Hybride) : Le système structure les Tweets (Sentiment, Priorité, Topic) via des LLM en intégrant une **stratégie de repli**. Le mode "Automatique" permet une bascule fluide vers un modèle local ou secondaire si l'API principale échoue, assurant la continuité d'activité.

1.4 Supervision et Contrôle d'Administration

L'interface utilisateur sert de centre de contrôle pour l'administrateur, offrant une flexibilité totale et des garanties de sécurité.

Gestion des Accès (Sécurité) :

Le système intègre une gestion complète des utilisateurs et des rôles. Cette distinction est cruciale : seuls les administrateurs peuvent configurer les modèles LLM et visualiser les logs techniques, tandis que les analystes accèdent uniquement aux tableaux de bord.

Orchestration Événementielle et Tâches :

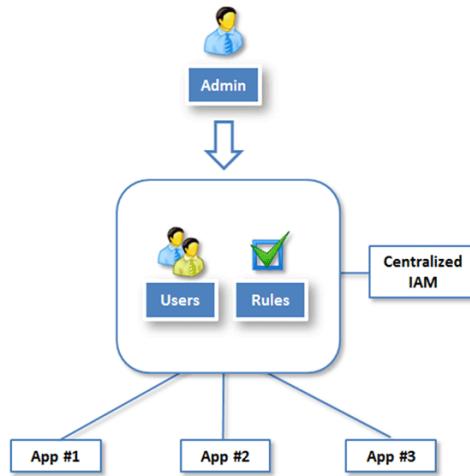


FIGURE 2 – Gestion des rôles

Email	Role	Created	Actions
admin@free.com	admin	10/11/2025	
user@free.com	data Analyst	18/11/2025	

FIGURE 3 – Interface de gestion des Utilisateurs, illustrant le contrôle d'accès basé sur les permissions.

Name	Description	Permissions	Actions
admin	Administrator role with all permissions	manage users, manage datasets, manage tasks, manage app, view overview, read users, read datasets, read tasks	
data Analyst		manage datasets, manage tasks, view overview, read analysis, read datasets, read tasks	
Product Manager		manage tasks, manage datasets, manage app, read datasets, view overview	

FIGURE 4 – Interface de gestion des Rôles, illustrant le contrôle d'accès basé sur les permissions.

Grâce à RabbitMQ, chaque étape du traitement émet des événements. Cette architecture événementielle permet une **supervision granulaire** de l'état du pipeline.

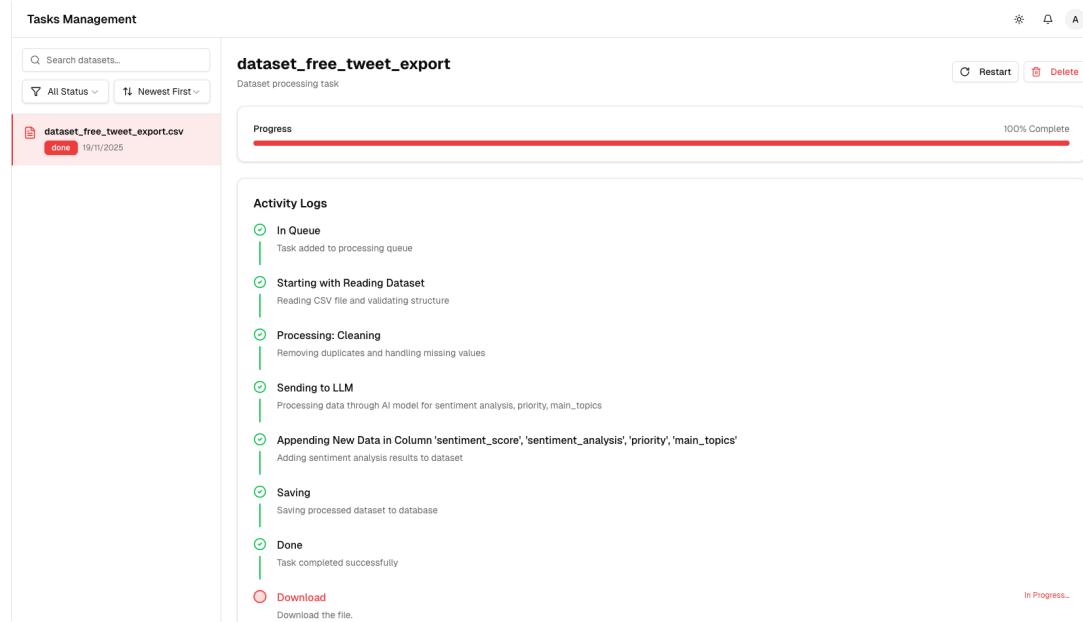


FIGURE 5 – Vue des Logs Système et Événements, fournissant une traçabilité détaillée des actions et des statuts de tâche.

Configuration Modulaire des Modèles IA :

L'administrateur peut adapter la solution aux contraintes externes (coût, confidentialité) en changeant dynamiquement l'API utilisée (OpenAI, Gemini, Mistral) ou en optant pour des modèles locaux (**Ollama**).

Add Local Model
Configure an Ollama local AI model

Model Name
e.g., Qwen3:4b

API Base URL
localhost:11434

API Key
Enter API key

Retry Requests 3
Max: 10

Paginate Rows Limit 500
Max: 5000

Add Model

FIGURE 6 – La page d'accueil de notre application

Add External Model
Configure a cloud-based AI service

Model Name
e.g., GPT-4

Provider
e.g., Google, OpenAI, UX Pilot AI

API Base URL
https://api.example.com

API Key
Enter API key

Retry Requests 3
Max: 10

Paginate Rows Limit 500
Max: 5000

Add Model

FIGURE 7 – La page d'accueil de notre application

2 Définition et Calcul des KPI SAV

Les indicateurs de performance clés (*Key Performance Indicators* - KPI) sont calculés dynamiquement dans le **frontend** sur la base des données enrichies.

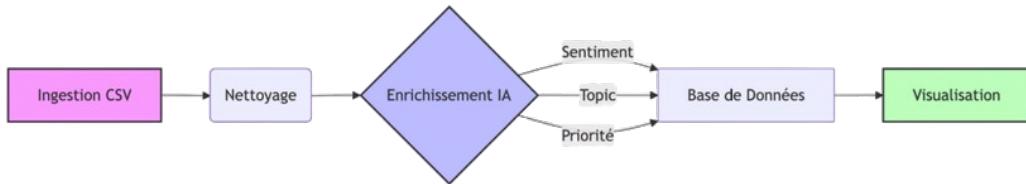


FIGURE 8 – Interface de Configuration des Modèles IA, illustrant la modularité pour le choix de l’API ou du modèle local.

2.1 KPI de Volume et de Flux

- **Volume Net** : Nombre de tweets pertinents après l’élimination du bruit.
- **Ratio de Bruit** : Pourcentage de tweets rejetés (doublons, spam) par rapport au volume initial.

2.2 KPI de Satisfaction et d’Urgence

- **Répartition des Sentiments** : Classification positive, neutral, negative post-analyse LLM.
- **Indice de Priorité** : Segmentation en High (Critique), Medium, Low basée sur des termes d’urgence identifiés.
- **Cartographie des Thèmes** : Identification du `main_topic` (*Réseau, Facture, Service Client*) pour le routage.

2.3 KPI d’Engagement et Viralité

- **Taux de Réponse (Reply Rate)** : Mesure de l’interactivité en analysant les métadonnées `in_reply_to`.
- **Viralité des Plaintes** : Corrélation entre le sentiment Négatif et le nombre de Retweets, indiquant l’amplification des sujets de mécontentement.

3 Exploration Visuelle et Interprétation

L'application génère des tableaux de bord interactifs pour la direction, transformant les données enrichies en *insights* opérationnels.

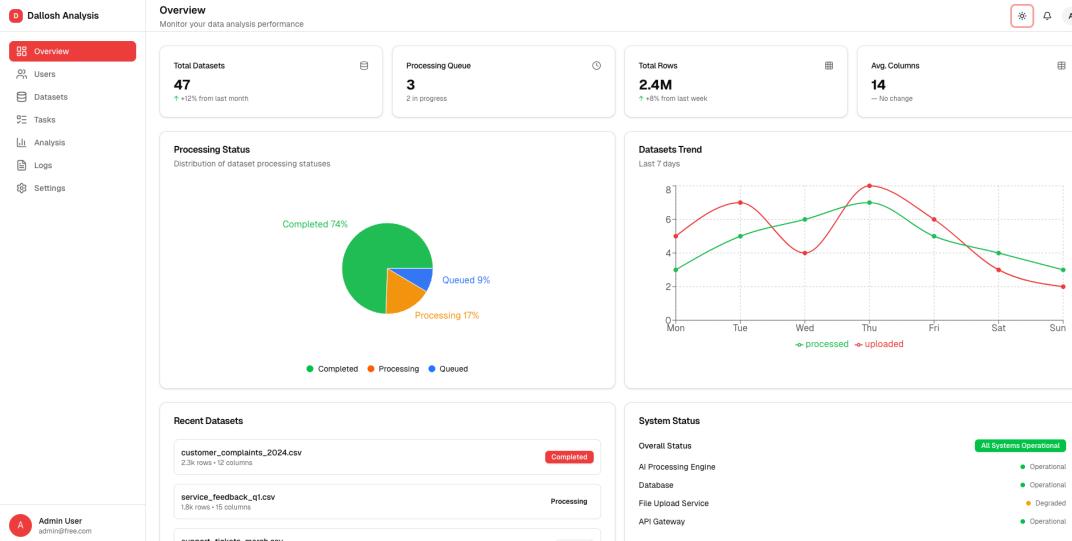


FIGURE 9 – Tableau de Bord Analytique Global, présentant une vue synthétique des KPI de performance du système et de satisfaction client.

3.1 Analyse Temporelle et Sentimentale

Les courbes d'évolution temporelle permettent une détection proactive des incidents. Une **hausse concomitante du volume et du sentiment négatif** signale un incident majeur (ex : panne réseau).

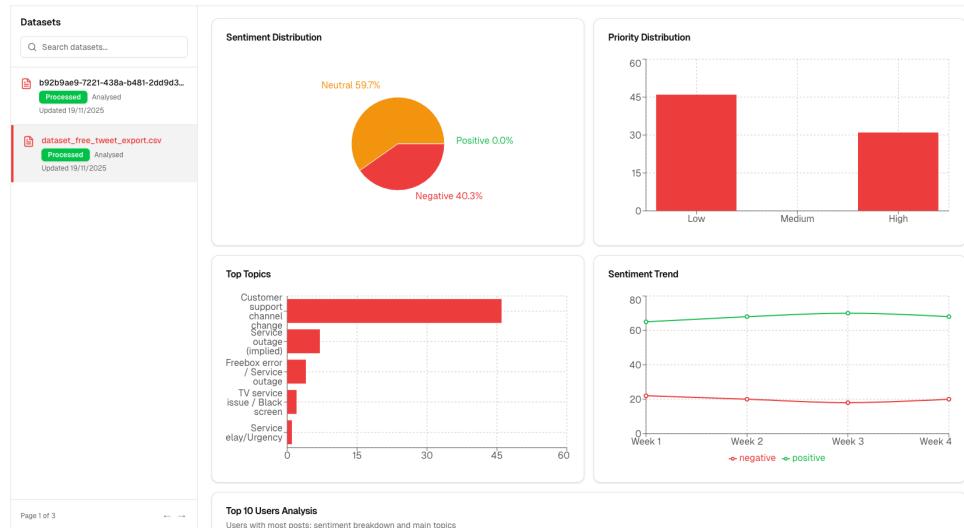


FIGURE 10 – Timeline de l'Évolution des Sentiments, outil clé pour l'identification des pics d'insatisfaction et leur corrélation temporelle.

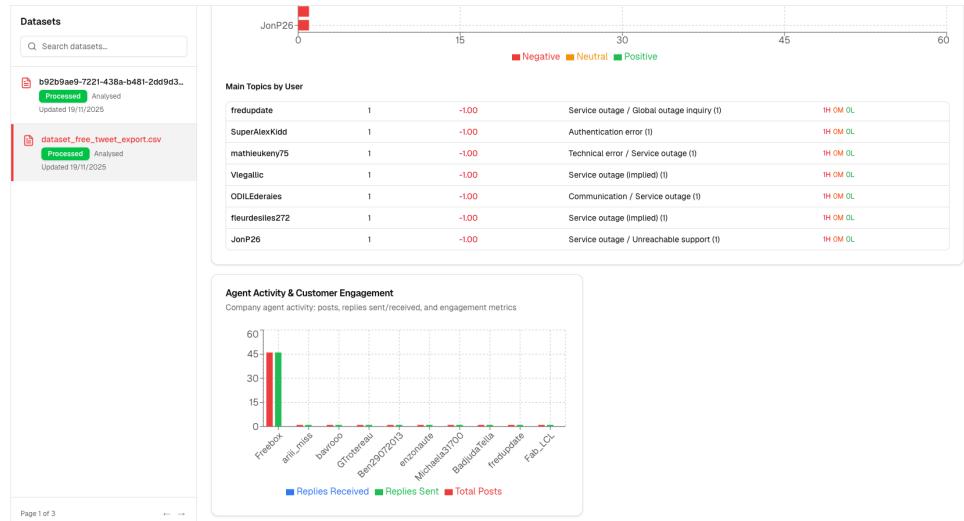


FIGURE 11 – Timeline de l’Évolution des Sentiments, outil clé pour l’identification des pics d’insatisfaction et leur corrélation temporelle.

3.2 Analyse Croisée (Topics vs Sentiments)

L’histogramme empilé fournit une cartographie des problèmes pour la prise de décision.

- Constat Critique :** Le sujet Réseau concentre le taux le plus élevé de négativité et de priorité haute.
- Constat Opérationnel :** Les sujets Administratifs maintiennent une répartition majoritairement neutre, indiquant une gestion stable de ces requêtes.

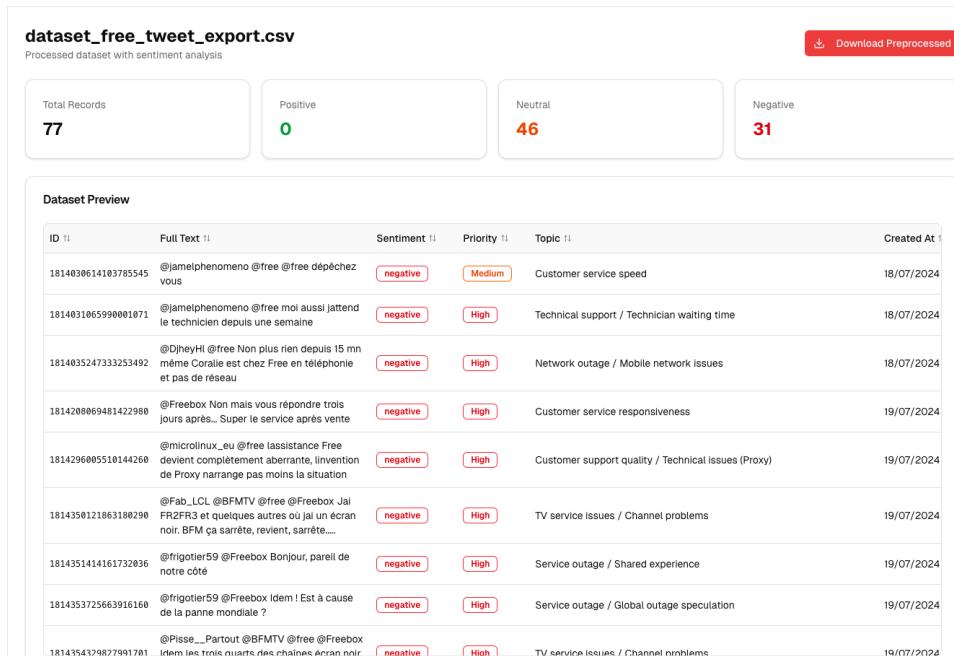


FIGURE 12 – Distribution croisée des Topics et Priorités, mettant en lumière les thèmes critiques nécessitant une intervention immédiate.

4 Qualité des Données et Limites du Modèle Actuel

L'analyse de la qualité montre que le processus de nettoyage est efficace, bien qu'il présente des limites inhérentes à la nature de la donnée :

- **Impact du Nettoyage :** L'élimination stricte des doublons et des spams de bots (*assainissement*) assainit la base pour l'IA mais peut sous-estimer l'ampleur d'un bad buzz artificiel.
- **Limite du Traitement par Lots :** Le traitement actuel par lots fixes (500 lignes) ne prend pas en compte la longueur variable des textes (tokens), ce qui peut engendrer des erreurs ou des troncatures si la limite contextuelle du LLM est atteinte. Ce point est adressé dans la section perspectives via la stratégie de *Smart Batching*.

5 Conclusion Technique

La solution *Dallosh* offre une infrastructure **résiliente** et **flexible**, capable de transformer des données brutes en *insights* visuels hautement exploitables pour la direction de Free.