
BIG DATA
ANALYSE DE SENTIMENT SUR
TWITTER

**Master en Intelligence Artificielle et Science
des Données**

Réalisé par:

MOHAMED AMINE FAKHRE-EDDINE

SAMADY AHMED

CHIBANI FAHD

Supervisé par:

Pr. Yasyn EL YUSUFI

19/05/2024

Contents

1	Architecture	3
1.1	Interface Utilisateur :	3
1.2	API (Backend) :	3
1.3	Système (Kafka et Spark) :	3
1.4	Docker :	4
1.5	Modèle :	4
2	Interface Utilisateur - Démonstration:	4
2.1	Interface Home	4
2.2	Interface prédiction text	5
2.3	Interface prédiction fichier	5
2.4	Interface d'historique de jobs	6
2.5	Interface de détails d'une	6
2.6	Graphe nombre de prédictions	7

1 Architecture

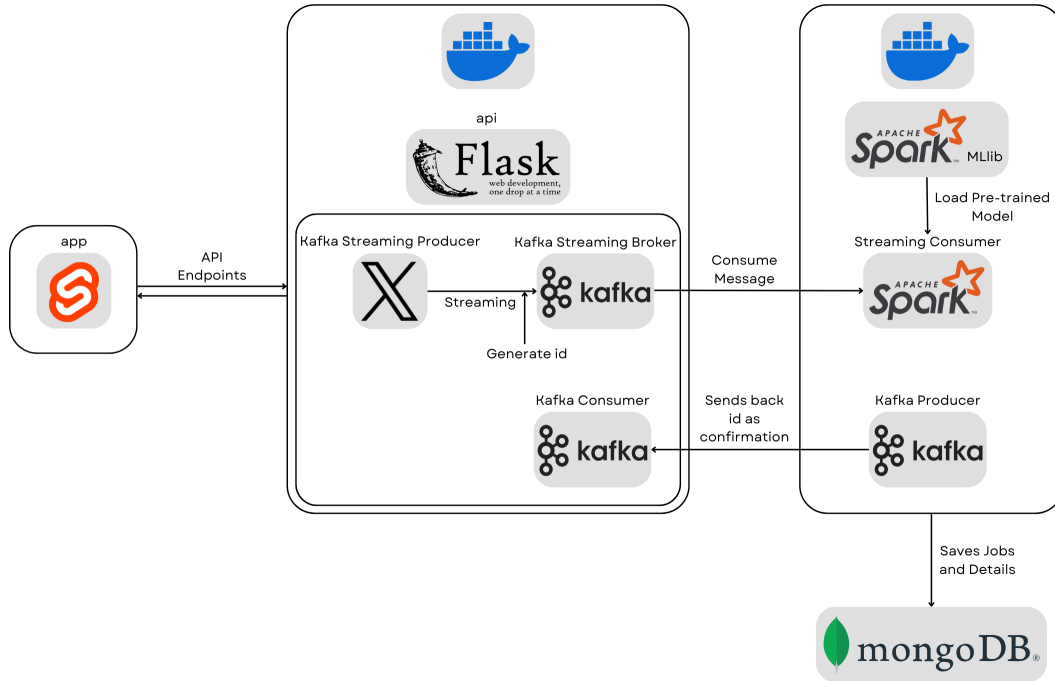


Figure 1: Architecture du système

Notre système est alimenté par un pipeline Kafka et Spark qui ingère des publications Twitter en temps réel. Les données sont traitées et analysées pour fournir des prédictions instantanées. L'architecture du système est la suivante :

1.1 Interface Utilisateur :

L'interface utilisateur est une application web développée avec Svelte qui permet aux utilisateurs d'interagir avec le système. Ils peuvent consulter l'historique des tâches et les détails, ainsi qu'initier de nouvelles tâches pour prédire le sentiment d'un seul tweet ou de plusieurs tweets contenus dans un fichier.

1.2 API (Backend) :

L'API est un service RESTful développé avec Flask qui sert d'intermédiaire entre l'interface utilisateur, le système et la base de données. Elle est responsable de la gestion des requêtes utilisateur, de leur traitement et du retour des résultats.

1.3 Système (Kafka et Spark) :

Le système est composé de deux composants principaux : le pipeline Kafka et le job Spark. Le pipeline Kafka ingère les publications Twitter et les envoie au job Spark pour traitement. Le job Spark pré-traite, analyse les données et fournit une prédiction de sentiment instantanée.

Après la prédiction, les résultats sont stockés dans une base de données MongoDB pour référence future. Ensuite, un ID est envoyé à l'interface utilisateur pour confirmation. L'utilisateur peut ensuite interroger la base de données pour obtenir les résultats de la tâche.

1.4 Docker :

Le système et l'API sont containerisés avec Docker pour assurer la portabilité et faciliter le déploiement. Les conteneurs sont orchestrés avec Docker Compose pour gérer les composants du système et leurs dépendances.

1.5 Modèle :

Le modèle d'analyse de sentiment est un modèle pré-entraîné de validation croisée et de régression logistique, entraîné sur un ensemble de données de plus de 63 000 tweets. Le modèle est utilisé pour prédire le sentiment d'un tweet comme étant soit positif, négatif, neutre ou non pertinent. Le modèle fonctionne avec une précision de 93

Cette architecture permet une analyse de sentiment efficace et en temps réel des tweets, en utilisant une combinaison puissante de technologies modernes pour l'ingestion, le traitement et l'analyse des données.

2 Interface Utilisateur - Démonstration:

2.1 Interface Home

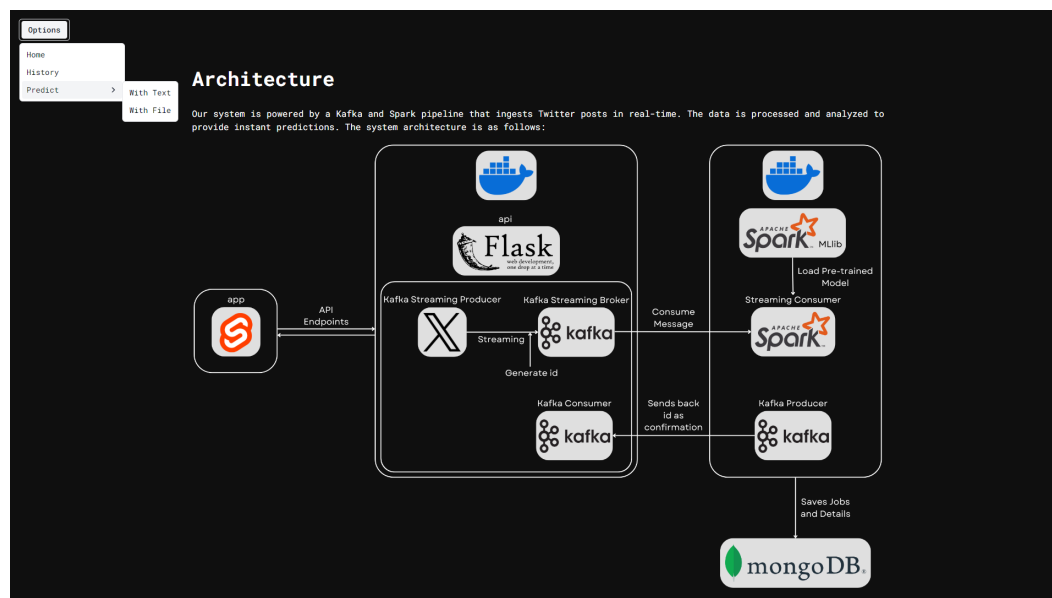


Figure 2: Interface Home

2.2 Interface prédiction text

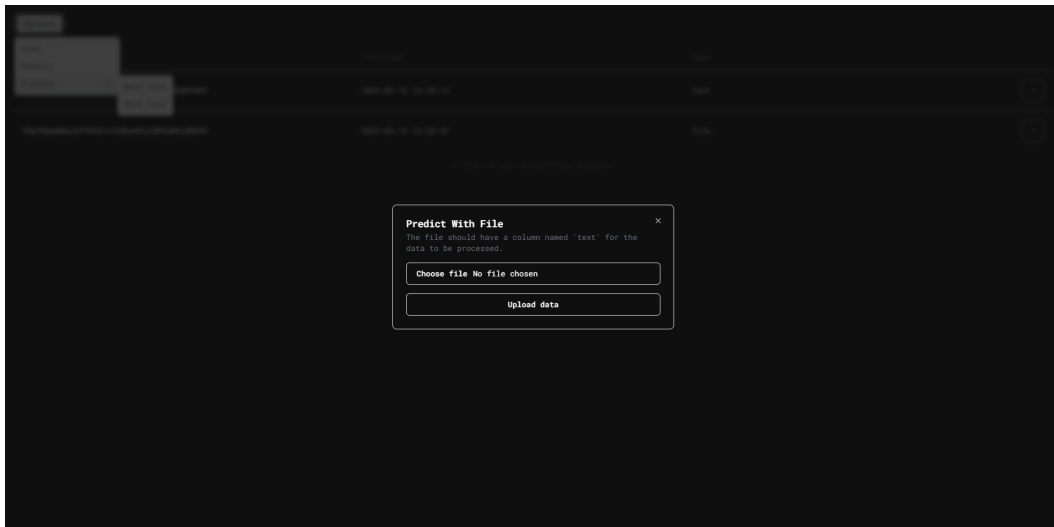


Figure 3: Interface prédiction text

2.3 Interface prédiction fichier

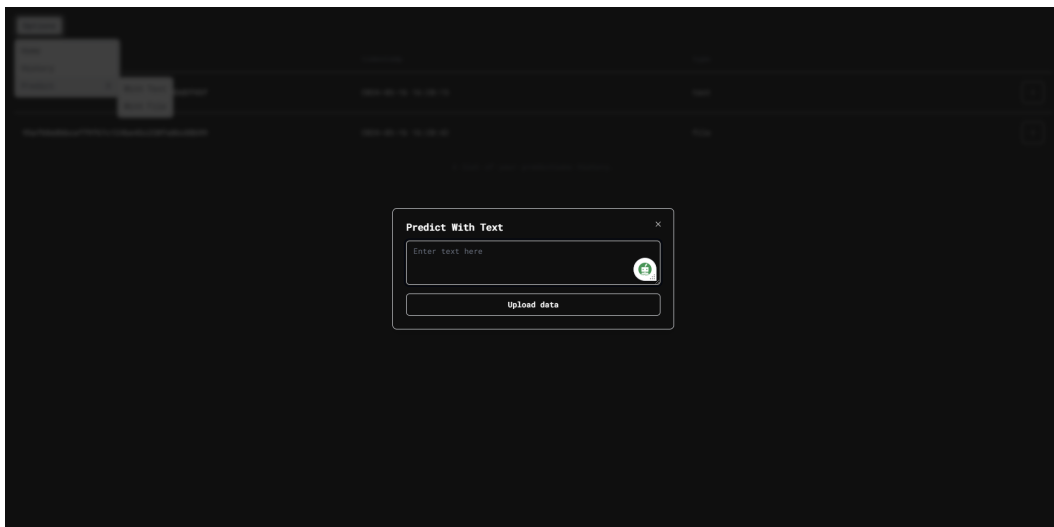


Figure 4: Interface prédiction fichier

2.4 Interface d'historique de jobs

Options

ID	timestamp	type
7f647da14e86f87b8dda81b84427575d3d5f45f	2024-05-16 16:28:15	text
95af68e8bbcaff9f67c124be45c238fa86c88099	2024-05-16 16:28:42	file

A list of your predictions history.

Figure 5: Interface d'historique de jobs

2.5 Interface de détails d'une

Options

Predictions count

text	probability	prediction
game sucks	Irrelevant: 0.02397956215633116 Negative: 0.9827977484878484 Neutral: 0.12626166719245482 Positive: 0.046961822163365965	Negative

A list of your predictions history.

Figure 6: Interface de détails d'une

2.6 Graphe nombre de prédictions

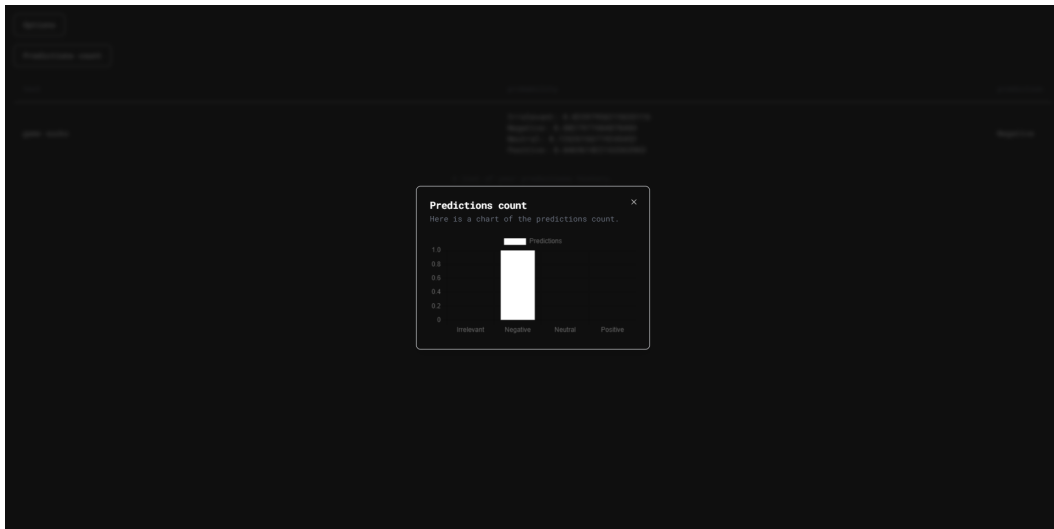


Figure 7: Graphe nombre de prédictions