

Projet d'Analyse des Données de Catastrophes : Traitement en Temps Réel et par Lot

Vue d'ensemble du projet :

I. Objectif :

L'objectif de ce projet est de mettre en œuvre à la fois le traitement par lots et le traitement en temps réel pour analyser et prédire les catastrophes, en mettant particulièrement l'accent sur les incendies de forêt. Vous exploiterez les données historiques des catastrophes pour l'analyse par lots et les données environnementales en temps réel pour le traitement en flux. Le projet impliquera à la fois l'analyse des données ainsi que l'utilisation du **machine learning** et du **deep learning**.

II. Portée :

● Traitement par lots :

Analyser les données historiques des catastrophes (incendies de forêts, tremblements de terre, tornades, tsunamis, etc.) afin d'identifier les tendances et les schémas.

● Traitement en temps réel :

Utiliser les données environnementales en temps réel (par exemple, température, humidité) pour prédire les occurrences d'incendies de forêts et évaluer leur risque, ainsi que des images satellites pour détecter la présence de feu.

Objectifs du projet :

I. Traitement par lots :

- Collecter et traiter les données historiques des catastrophes, en se concentrant sur les incendies de forêt, les tremblements de terre et autres catastrophes.
- Analyser les tendances basées sur des métriques clés comme les dommages, la localisation et le temps.
- Présenter les résultats via Power BI avec des visualisations interactives et des tableaux de bord.

II. Traitement en temps réel :

- Mettre en place des pipelines de données en temps réel pour capturer les données environnementales (par exemple, température, humidité) liées aux risques d'incendie de forêt.
 - Implémenter un modèle de machine learning et de deep learning pour prédire la probabilité des occurrences d'incendies de forêt basées sur ces données en temps réel.
-

Tools & Technologies :

- Programming Languages: Python
- Data Analysis and Visualization: Pandas, Power BI
- Stream Processing: Apache Kafka, Apache Spark Streaming
- Machine Learning: Scikit-learn, TensorFlow or PyTorch for predictive models
- APIs: Weather and environmental data from sources like NASA, NOAA, and other weather APIs
- Data Integration & Workflow: Talend, Apache Airflow
- Data Storage: Data Lake, Data Warehouse, MongoDB or Cassandra for storage

Sources de Données :

I. Traitement par Lots :

- **Données mondiales sur les catastrophes** : Datasets provenant de :

- a. [Kaggle - Données mondiales sur les catastrophes](#)

Example:

1. <https://www.kaggle.com/datasets/mexwell/significant-volcanic-eruption-database>
2. <https://www.kaggle.com/datasets/harshalhonde/tsunami-events-dataset-1900-present>
3. <https://www.kaggle.com/datasets/dani215/fire-dataset>
4. <https://www.kaggle.com/datasets/ananthu017/california-wildfire-incidents-20132020>
5. <https://www.kaggle.com/datasets/headsortails/us-natural-disaster-declarations>
6. <https://www.kaggle.com/datasets/kazushidachi/global-landslide-data>
7. <https://www.kaggle.com/datasets/usamabuttar/significant-earthquakes>
8. <https://www.kaggle.com/datasets/andrewmvd/tsunami-dataset>
9. <https://www.kaggle.com/datasets/pedroclapers/near-earth-objects-1900-to-2200>

- b. [NOAA - National Oceanic and Atmospheric Administration](#)

- c. [EM-DAT - Centre pour la recherche sur l'épidémiologie des catastrophes](#)

- **Données spécifiques sur les catastrophes** : Par exemple, incendies de forêt, tremblements de terre, etc., avec des colonnes pertinentes comme :

- ✓ Localisation
- ✓ Heure
- ✓ Dégâts
- ✓ Victimes

II. Traitement en Temps Réel :

- **Données environnementales et en temps réel**: Utilisation des API suivantes pour obtenir des données relatives aux incendies de forêt :

- a. [OpenWeather](#)

Livrables Clés :

I. Traitement par Lots :

- Une analyse détaillée des données historiques sur les catastrophes, avec des aperçus sur les tendances, les régions et les périodes pour différents types de catastrophes (par exemple, incendies de forêts, tremblements de terre, tsunamis, etc.).
- Des rapports Power BI avec des tableaux de bord interactifs pour visualiser les principales métriques et tendances des données sur les catastrophes.

II. Traitement en Temps Réel :

- Un pipeline de données en temps réel pour traiter les données environnementales, y compris la température, l'humidité et la vitesse du vent, afin d'évaluer les risques d'incendie de forêt.
- Un modèle d'apprentissage machine pour prédire les risques d'incendie de forêt en fonction des facteurs environnementaux en temps réel.
- Un modèle d'apprentissage profond pour analyser les images satellites afin de détecter la présence d'incendies de forêt.
- Un tableau de bord en direct intégrant les modèles d'apprentissage machine et d'apprentissage profond pour visualiser et détecter les risques d'incendie de forêt en temps réel.

Contraintes et Risques :

- **Disponibilité des données** : Vérifier que les sources de données sont fiables, complètes et accessibles pour le traitement par lots et en temps réel.
 - **Limites des API** : Vérifier les limites d'utilisation et les erreurs possibles des données des API externes.
 - **Performance du modèle** : S'assurer que le modèle de machine learning donne des prédictions fiables malgré les variations environnementales.
 - **Contraintes de temps** : Gérer efficacement le calendrier du projet pour intégrer à la fois le traitement par lots et en temps réel dans le délai imparti.
-

Budget :

Le budget du projet est en cours d'évaluation et sera communiqué une fois les détails finalisés.

Calendrier du Projet :

Date limite de soumission : 30 avril 2024, 23h59