

Livrable 3 : Consommation des données

Introduction

Ce projet a pour objectif d'utiliser le portail développeur du journal américain NY Times, qui propose plusieurs API à explorer et de créer une application fonctionnelle qui valorise les données récoltées par ses API.

Ce projet s'articule autour de plusieurs étapes pour lesquelles un livrable devra être rendu. Ces étapes sont :

- Découverte des sources de données disponible :
- Organisation des données
- Consommation des données
- Déploiement
- Automatisation des flux

Nous avons pour problématique d'estimer la popularité d'un sujet sur le New York Times et sur les ventes de livres

L'objectif de ce livrable est de consommer nos données récoltées, c'est à dire de les valoriser. Pour cela nous pouvons réaliser un modèle de machine learning et ou un tableau de bord

Dash

Compte tenu de notre problématique et de notre projet, nous avons décidé de réaliser un Dashboard. En effet, nous avons des données qui ne se prêtent pas au machine learning puisque ce sont des données textuelles comme des mots-clés, de très courts paragraphes sur lesquels il est difficile de prédire ou de déterminer quelques choses. Cependant, ce type de données est idéal pour des statistiques simples sur une échelle de temps comme le nombre d'occurrences par exemple.

Nous créerons donc un Dashboard afin de visualiser ces statistiques simples. Pour cela nous utiliserons l'outil Dash pour le créer.

A l'aide d'un script python, nous allons implémenter les différents graphs que nous souhaitons afficher, et grâce à un système d'input, l'utilisateur pourra choisir les mots-clés sur lesquels il souhaite obtenir des informations. Ce même script python va appeler notre API Elasticsearch afin de récupérer les données nécessaires.

Conclusion et Perspectives

Pour ce livrable, nous souhaitons que notre application soit fonctionnelle. Nous avons pour l'instant une seule fonctionnalité avec un type de requête vers Elasticsearch. Les perspectives à court terme sont d'augmenter le nombre de fonctionnalités et d'améliorer l'architecture de l'application dans le but de l'automatiser. Les perspectives à plus long terme seront de conteneuriser cette application.