|  |
| --- |
|  |
| Million Song Dataset |
| Projet d’Apprentissage Automatique - Centrale |

|  |
| --- |
| Adlen Afane;Nicolas Schmidt  07 mars 2013 |

Contenu

[I. Objectif du projet 2](#_Toc350379836)

[II. Analyse des données 2](#_Toc350379837)

[III. Fonctionnement de l’outil 3](#_Toc350379838)

[IV. Difficulté rencontrées 3](#_Toc350379839)

[V. Analyse des résultats 3](#_Toc350379840)

[Bibliographie 4](#_Toc350379841)

[Annexes 5](#_Toc350379842)

[Screenshot des différents écrans 5](#_Toc350379843)

# Objectif du projet

Nous avons choisi d’utiliser d’effectuer de la recommandation de chansons en nous basant sur les données du Million Song Dataset. Pour ce faire, nous avons utilisé 2 techniques pour trouver des chansons similaires :

* Cluster de chansons
* **Calcul des k plus proches voisins**

L’objectif est de pouvoir donner à l’outil une chanson et obtenir en sortie des chansons similaires à celle-ci. L’utilisation pratique peut être comme sur Spotify ou Amazon de proposer une recommandation pour la chanson / l’article que l’on a sélectionné, on peut également envisager la création automatique de playlist à partir d’une chanson.

# Analyse des données

Insérer les stats qu’on a sur le #de chansons, #artistes, # moyen de chansons par artistes…

# Fonctionnement de l’outil

Nous avons choisi de développer une interface web pour manipuler les données et générer nos calculs et sorties. Nous avons fait ce choix car cela permet une flexibilité plus « humaine » dans le choix des données, des algorithmes et des tests ainsi que pour la représentation des résultats et également pour s’approcher un peu plus d’un produit fini. Le fonctionnement et les possibilités de cet outil sont décrits dans le schéma suivant :



# Difficulté rencontrées

* Quels éléments d’entrée choisir ? -> Interface simple pour sélectionner et générer les données
* Quels algos choisir ? -> Interface simple pour choisir et tester ses algos
* **Evaluer la pertinence des recos / du clustering ? Représentation des résultats, Utilisation de l’API echonest pour tester des chansons**
* Passage à bcp de données (on teste avec 100)
* Données vides

# Analyse des résultats

* Capture d’écran des étapes
* **Besoin d’évaluations quantitatives du clustering !**

# Bibliographie

**Aucune source spécifiée dans le document actif.**

# Annexes

## Screenshot des différents écrans