



Haute Ecole Spécialisée
de Suisse occidentale

BDA

Rapport : Recommending Music and the Audioscrobbler Data Set

Rapport

Version 1.0, 29.05.2020

Samuel Torche

Ayrton Dumas

Marco Mattei

Groupe B



MASTER OF SCIENCE
IN ENGINEERING

Table des matières

1.Introduction	3
2.Description du dataset	3
3.Description des features / pre-processing afin d'extraire des features additionnelles	3
4.Questions d'analyse	3
5.Algorithmes appliqués	3
6.Optimisations réalisées	3
7.Tests et évaluations	3
8.Résultats obtenus	3
9.Améliorations futures	3
10.Conclusion	3

1.Introduction

2.Description du dataset

size, information it contains

Le dataset utilisé est appelé AudioScrobbler. Il s'agit d'une archive compressée contenant plusieurs fichiers texte. L'archive fait 135 MB compressée et 500 MB une fois décompressée.

User artist data

Le fichier user_artist_data.txt contient les utilisateurs et les artistes qu'il écoute. Chaque ligne correspond contient l'id de l'utilisateur, l'id de l'artiste et le nombre de fois qu'il a écouté une chanson de cet artiste, donc pour chaque artiste écouté par un utilisateur, il y aura une nouvelle ligne. Il y a 140'000 utilisateurs qui ont écouté en tout 24'296'858 artistes. Ce fichier contient donc 3 colonnes: userid, artistid, playcount

Artist data

Le fichier artist_data.txt contient les mappings artistes id -> nom de l'artiste. Ce fichier contient donc 2 colonnes: artistid, artist_name. Il y a 1'848'707 artistes dans ce fichier.

Artist alias

Le fichier artist_alias.txt contient les noms d'artiste qui sont souvent mal orthographiés ou qui sont sujets à variante (Depeche Mode mal orthographié en Depeche Mood par exemple). Ce fichier contient donc 2 colonnes: badid, goodid. Il y a 193'027 alias.

3.Description des features / pre-processing afin d'extraire des features additionnelles

4.Questions d'analyse

1. Peut-on obtenir des résultats intéressants en utilisant des technique de Market Basket Analysis telles que les règles d'association ?
2. Peut-on déduire les genres des musiques à l'aide d'un clustering des utilisateurs et leurs écoutes ?
3. Est-il possible de recommander non pas des artistes mais d'autres utilisateurs qui partagent les même goût avec ALS ?

5.Algorithmes appliqués

6.Optimisations réalisées

7.Tests et évaluations

8.Résultats obtenus

9.Améliorations futures

10.Conclusion