



**POLYTECHNIQUE
MONTRÉAL**

UNIVERSITÉ
D'INGÉNIERIE

DÉPARTEMENT DE GÉNIE INFORMATIQUE
École Polytechnique de Montréal

INF8245E : MACHINE LEARNING

Kaggle project

Auteurs :

GOURION Andréa

Matricule XXXXXXXX

TREMBLY Timothée

Matricule XXXXXXXX

ROHMAN Shahriar

Matricule 2164342

21 NOVEMBRE 2022

1 Introduction

2 Pre-processing

2.1 Problématique

Données manquantes, données en chaînes de caractères, données aberrantes, ect...

2.2 Data cleaning

On a enlevé les NaN. We simply removed all the rows that contained NaN values. We also removed the rows that contained the string "Insert smth"...

2.3 Data transformation

On a converti les string en int à partir de dictionnaires. Pour la colonne "Id", on a simplement enlevé "ConcertExperience_" devant l'Id du concert afin d'éviter de créer un dictionnaire trop gros et d'aller plus vite dans le traitement.

3 Modélisation

3.1 Comparaison des modèles

On a comparé les modèles suivants : SVM, KNN, Random Forest, Decision Tree, Logistic Regression, Naive Bayes, ect...

3.2 Modèle choisi

On a choisi le modèle XX qui donne les résultats suivants :