Projet : Rapport 1

GLO-4027/GLO-7027

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Jean Le Berre

Gabriel Philippon

Laura Pedenaud

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

16/02/2022

# Introduction

Le but de ce projet est de prédire le choix de vote des personnes ayant répondu au questionnaire en ligne du CES pour les élections fédérales canadiennes de 2019. Les données ont été fourni par ce même organisme. Les réponses aux questions en relation avec le choix du vote ont été supprimée pour 10% des participants. C’est à nous de prédire la position politique de ces personnes en utilisant les réponses des autres questions du sondage.

# Analyse globale des données

Le rapport de l’enquête nous dit qu’un pré-nettoyage a déjà été effectué. Ainsi, les réponses dupliquées, les questionnaires répondus trop vite ont déjà été supprimées. Un deuxième nettoyage est effectué supprimer les 3% de personnes qui n’ont rien indiqué dans les champs cps19\_votechoice ou cps19\_votechoice\_pr ou cps19\_vote\_unlikely ou cps19\_vote\_unlike\_pr ou cps19\_v\_advance.

Ensuite, tout le monde n’est pas soumis au même questionnaires. Les questions diffèrent suivant la nationalité, la province mais aussi de manière aléatoire. Cela sera à prendre en

On remarque que beaucoup d’attributs ont des valeurs manquantes (que 31% de valeur utile en moyenne sur les 621 attributs). Cependant, beaucoup d’attribut sont complémentaires, ils répondent à une partie de la question (Ex : un attribut par parti pour une question).

On a choisi de