

# Description du projet

*Stéphane Caron*

*2018-01-01*

## Abstract

Ce document a pour but d'expliquer le projet travaillé par trois joueurs de l'équipe de hockey des Dynamiques du Cégep de Sainte-Foy. Le projet consiste en une brève introduction aux statistiques, à la programmation et aux techniques d'apprentissage automatique permettant de résoudre des problèmes concrets.

## Contents

Mise en contexte . . . . .	1
Description du projet . . . . .	1
Description des données . . . . .	2
Méthodologie . . . . .	2
Analyse des résultats . . . . .	2

## Mise en contexte

Ce projet a été réalisé dans le cadre du programme de “tutorat”, mis en place par Christian Larue, entraîneur des Dynamiques du Cégep Sainte-Foy. Le programme a pour but de présenter aux joueurs actuels de l'équipe différents domaines dans lesquels certains anciens joueurs oeuvrent actuellement.

Ce projet spécifique permet de donner une brève introduction aux domaines des mathématiques et statistiques, en plus de toucher à plusieurs concepts en lien avec la programmation et l'analyse de données. Ces concepts peuvent s'appliquer à plusieurs autres domaines, notamment l'informatique et l'actuariat. Pour plus d'informations sur ces domaines en particulier, voici quelques liens pertinents:

Mathématiques et statistiques:

- Département de mathématiques et statistique de l'Université Laval
- Data science and statistics jobs

Informatique et programmation:

- Département d'informatique et génie logiciel de l'Université Laval
- McGill School of Computer Science
- Data science and analytics in sports

Actuariat:

- École d'actuariat de l'Université Laval
- Society of actuaries
- Casualty actuarial society

## Description du projet

Comme mentionné dans la section précédente, ce projet consiste en une brève introduction à certaines méthodes statistiques s'appliquant à des problèmes concrets. Le concept principal introduit dans ce projet s'appelle le “clustering”. Le *clustering* est une méthode statistique permettant de regrouper des données dans différents groupes partageant des caractéristiques similaires. Il existe plusieurs méthodes de clustering,

basées sur différents algorithmes, qui permettent d'obtenir différents résultats dépendamment du contexte. Ces méthodes sont fréquemment utilisées dans plusieurs contextes dans ces différents domaines:

- Marketing:
  - Pour la segmentation de marché et l'analyse de prospects potentiels.
  - Pour la rétention de clients actuels.
  - Pour l'analyse géographique de marchés potentiels.
- Finance:
  - Regroupement d'actions présentant des caractéristiques similaires (gestion de portefeuille).
  - Établissement de caractéristiques pour identifier de potentiels non-payeurs.
- Médecine:
  - Recherche de caractéristiques présents chez un type de patient.

Dans ce projet, nous utiliserons une méthode de clustering précise, la méthode k-means (qui est décrite un peu plus loin), pour établir des styles de joueurs de hockey. En effet, nous utiliserons des statistiques de joueurs de hockey de la LNH (section suivante) pour établir des groupes de joueurs partageant des caractéristiques similaires selon certaines statistiques.

## **Description des données**

### **Jeu de données**

### **Nettoyage des données**

## **Méthodologie**

### **Méthode k-means**

## **Analyse des résultats**