Report V0

March 31, 2022

Contents

1 Context

Motivation is an essential parameter for the engagement of young people in voluntary physical activity or sport. Traditionally, the level of motivation is measured by questionnaires which are restrictive to fill in. In connection with the University of Lille, an application has been developed to identify the type of motivation on a more functional mode. The principle consists in asking the student 1 question: "In PE, what was the sport that you enjoyed the most?" Then we would indicate to him: "We are now going to present you with words that will allow you to describe your feelings about this sport. Your job is to indicate, as quickly as possible, whether you agree or disagree with these proposals by clicking yes or no. The response time was taken into account in each answer. If this time is short, it means that the term seems obvious. For example, if the sport is "soccer", the student could answer "yes" quickly to the qualifier "fun", "no" quickly to the qualifier "beauty".

2 Description of data

We have the results of a questionnaire carried out by 1070 participants with the time in milliseconds of the answers taken into account. The possible answers to each question are "yes", "no", "I don't know". When the answer to a question is "yes" then the time value is positive and negative in the case of "no", zero in the case of "I don't know".

3 Objective

3.1 Main Objectives

The objective of this work is to identify profiles of practitioners based on positive or negative qualifiers, i.e., to assign a profile to each cluster of the data and to estimate the strength

of these profiles, i.e., the number of clusters or profiles that are most representative of the data.

3.2   Specific Objectives

We will first perform a pre-processing of the data by renormalizing the data (by row), removing outliers, completing or removing missing values, and then to analyze the data we will use different algorithms such as: K-means, principal component analysis, decision trees. Finally we will compare the different algorithms by intermediate validations.