

Projet Cogmont

Azat, Lucas, Matthieu, Etienne

March 13, 2019

Contents

Introduction	1
Contexte / Sujet	1
Les besoins d'un changement éducatif	2
Les pedagogies	2
Présentation des données de départ	2
Importation des données	2
Nettoyage des données	3
Méthodologie	3
Analyse factorielle	3
Regression Logistique	3
Classification Ascendante Hiérachique	3
Arbre de regression	3
Analyse exploratoire	4
Univariée	4
Multivariée	4
Recherche de regroupement de variables (cah sur les variables)	4
Recherche de différence significative entre p1 et p2	4
Regression logistique	4
Tests sur les différentes réponses	4
Arbre de regression	4
Conclusion	4

Introduction

Contexte / Sujet

Ce projet nous à été proposé par l'Institut des sciences cognitives - Marc Jeannerod spécialisé dans la neuroscience. L'UMR 5304 créée en 2007 est un des deux laboratoires de l'Institut des Sciences Cognitives

– Marc Jeannerod. L'UMR 5304 est un laboratoire interdisciplinaire qui intègre l'expertise de chercheur des Sciences de la Vie (psychologie cognitive, neurosciences) et de médecine (pédo-psychiatrie, neuro-pédiatrie) avec celle de chercheur des Sciences Humaines et Sociales (linguistique computationnelle et théorique et philosophie) pour étudier la nature et la spécificité de l'esprit humain.

Les besoins d'un changement éducatif

Le député et mathématicien Cédric Villani a publié un rapport pour renforcer l'apprentissage des mathématiques à l'école. Les élèves français sont aujourd'hui plus que médiocres dans cette discipline. Pourtant, jusqu'en 1985, l'enseignement des maths en France était reconnu comme l'un des meilleurs. Or, en décembre 2016, dernière édition de classement Pisa (Programme for international student assessment) la France a fini 24ème sur 72, en recule par rapport à la dernière édition. Pour mettre un terme à cette tendance inquiétante de la dégradation du niveau des élèves français en mathématiques, le gouvernement est à la recherche de nouvel pédagogie d'enseignement des mathématiques.

Les pédagogies

Pédagogie Montessori: La pédagogie Montessori est une méthode d'éducation créée en 1907 par Maria Montessori.

La pédagogie se base sur trois principe:

-l'autodiscipline: les enfants sont libres de choisir l'activité qu'ils souhaitent faire parmi celles qui leur sont proposées.

-L'action en périphérique: Selon Maria Montessori, il est plus profitable d'agir sur son environnement plutôt que sur l'enfant lui-même (comme des classes multi-âge).

Pédagogie "Traditionnelle": La pédagogie traditionnelle est celle du modèle transmissif. Selon le triangle pédagogique de Jean Houssaye, cette pédagogie privilégie la relation entre l'enseignant et le savoir. Autrement dit, l'enseignant expose un savoir sous forme de cours magistral, généralement suivi d'exercices ou/et de leçons à apprendre. L'élève doit intégrer et appliquer le savoir exposé par l'enseignant.

Présentation des données de départ

Importation des données

Notre jeu de données est composé de trois fichiers Excel (.xlsx), avec les résultats de chaque promotion au test cognitif européen.

MathsJetons_2015-2016.xlsx : Pour l'année 2015/2016.

MathsJetons_2015-2016.xlsx : Pour l'année 2016/2017.

MathsJetons_2016-2017.xlsx : Pour l'année 2017/2018.

Chaque jeu de données représente les résultats questions par questions (en comptant les sous-questions) des élèves ainsi que leurs catégories pédagogiques et des informations telles que l'encadrant, le niveau scolaire, la langue natale, l'âge, le type de classe (mélangé entre plusieurs section ou pas), l'année de passage du test et leur école. Il y a 10 questions divisées en sous questions, ce qui fait un total de 34 réponses. Chaque question est indépendante et pour répondre à la sous question suivante il faut une bonne réponse à la sous-question précédente, sauf pour la question 4, toutes ses sous-questions sont indépendantes. Une bonne réponse correspond à un 1 et une mauvaise réponse à un 0, sauf pour la réponse à la question 1 qui est la valeur de comptage maximale de l'enfant. Ici les élèves viennent tous de l'école Ambroise Croizat à Vault-en-Velin.

Nettoyage des données

Les données ayant déjà été travaillées l'année dernière le travail nécessaire en datamanagement n'a pas été excessif. Il nous a fallu tout de même renommer certaines variables pour les rendre plus lisibles, supprimer certaines questions car elles n'avaient été posées qu'à certaines classes... Les questions étant posées de manière à ce qu'au sein d'une même tâche, il faille réussir les questions dans l'ordre pour passer à celles plus dures nous avons beaucoup de NA dès qu'une question est dure. Nous avons donc décidé de changer ces NA et les considérer comme une question que l'élève n'aurait pas réussi. Et pour ne pas passer à côté de l'information : "n'a pu aller plus loin", nous avons créé deux jeux de données : un vectoriel, un composé de scores correspondant à la somme des questions au sein d'une même tâche. ## Recodage des variables

Méthodologie

Afin de répondre au mieux à notre problématique nous avons fait le choix d'utiliser plusieurs méthodes statistiques différentes pour analyser nos données. Pour cela nous avons dans un premier temps utilisé une méthode qui permet de résumer l'information globale du jeu de données : l'analyse factorielle, et la classification ascendante hiérarchique (pour faire des regroupement de variables). Puis dans un but prédictif nous avons utilisé la régression logistique et les arbres de régressions. Plusieurs tests ont été fait en parallèles, comme celui du χ^2 , de student..

Analyse factorielle

La méthode d'analyse factorielle que nous avons utilisé ici est l'analyse des correspondances multiples (ACM), qui est une méthode de synthétisations du nombre de dimensions pour les données qualitative. Cela nous permet d'appréhender plus rapidement le jeu de donnée, et avoir une première idée de ce qui diffèrent les individus entre eux (ou ce qui les rapproche). L'ACM permet dans un nuage à N dimension, en cherchant les plans orthogonaux qui maximisent la variance entre les individus, à résumer celles ci en 4 voir 5 dimensions. L'ACM a été réalisée sur les réponses aux questions en variables actives (celles qui définissent le placement des individus sur le graphe) avec les variables portant sur la pédagogie, la question 1, et l'âge en illustratives (ajoutée après le placement des individus sur le graphe).

Regression Logistique

Notre problématique étant de voir s'il existe un lien entre la façon d'enseigner et les réponses au test, nous avons voulu essayer de prédire la méthode d'enseignement à l'aide des réponses des élèves avec la régression logistique. Cette méthode permet de modéliser une classification, à l'aide notamment de l'odds ratio

Classification Ascendante Hiérarchique

N'ayant aucune information au préalable sur le thème des questions, leur regroupement... etc Mais sachant que certaines questions faisaient appel aux mêmes compétences. Nous avons utilisé une variante de la classification ascendante hiérarchique (CAH) afin de partitionner nos variables.

Arbre de regression

L'arbre de régression est une technique d'apprentissage supervisé, qui permet en analysant un grand nombre de données, de prédire une variable à expliquer. Ils sont beaucoup utilisés dans le domaine du marketing, et plus récemment dans le domaine du machine learning (apprentissage automatique). Dans un premier temps il s'agit d'exprimer la variable à expliquer en fonction d'un maximum de variables explicative, puis d'élaguer

l'arbre afin de minimiser l'erreur, soit l'écart entre la valeur prédite et la valeur réelle. Cela revient donc à faire une régression logistique sur les données, puis d'appliquer l'algorithme de construction d'arbre à partir des résultats.

Analyse exploratoire

Univariée

Multivariée

Recherche de regroupement de variables (cah sur les variables)

Recherche de différence significative entre p1 et p2

Regression logistique

Tests sur les différentes réponses

Arbre de regression

Conclusion