

Solucion Datathon

Sahlre

1 de julio de 2017

Introducción

Los datos abiertos relacionados al sector educativo contienen diversas mediciones (variables o características) asociadas a los hogares que podrían dar solución al desafío de inversión educativa en Colombia. Una de ellas es el nivel educativo de las personas, variable que se convierte en la variable resultado o dependiente que se intentará predecir con las otras características asociadas a cada una de los individuos, como por ejemplo, sólo por mencionar algunas, el género, el material predominante de las paredes exteriores y estado laboral.

Con dichas características se construyeron diversos algoritmos supervisados para predecir el nivel educativo de las personas: *Uno vs todos Regresión Logística Multivariada, Árbol de clasificación, Bagging con árboles de clasificación, Boosting con árboles de clasificación y Bosques Aleatorios de clasificación*. La decisión de incluir varios algoritmos es poder propener en esta solución ambos aspectos: predicción e interpretabilidad. Es sabido el *trade off* que existe entre esos aspectos, a mayor capacidad de predicción de un algoritmo menor será su interpretabilidad.

Los principales resultados son: XXXXXXXXXXXXXXXX.

Este documento se organiza de la siguiente manera: XXXXXXXXXXXXXXXX.