

Predicting Schools Performance

Introduction to Data Science Bootcamp

Fundação Getulio Vargas - July 2018

This contest aims to predict how a high schools will perform in the Brazilian national exam ENEM. We are making available a number of datasets containing information about schools in São Paulo city. The goal is to use information from the different sources in order to predict the performance of the students in the ENEM exam in 2015. In particular, the file [ENEM2015.csv](#) contains the classification of the high schools in ENEM 2015, classifying each school in a scale from 0 to 4, where 4 corresponds to the best schools.

In order to build a classification model each team should:

1. Extract from each data set information related to the schools listed in the file [ENEM2015.csv](#). Each school has a unique identification number, which must be used to find the school in each data set.
2. Build a new data set where the columns are a subset of the columns from the given data set.
3. Use the built data set to perform training and test.

The team with the smallest test error will win the contest.