

# Mineração de dados abertos-ENEM 2018

Alessandro Aparecido Barcellos<sup>1</sup>, Carlos Diego Nascimento Damasceno<sup>2</sup>, Prof. Dr. Seiji Isotani<sup>2</sup>

<sup>1</sup> Afiliação atual, e.g. Secretaria de Educação ou Pós-Graduação EAD em Computação Aplicada à Educação (ICMC/USP)

<sup>2</sup> Universidade de São Paulo (USP)

## INTRODUÇÃO

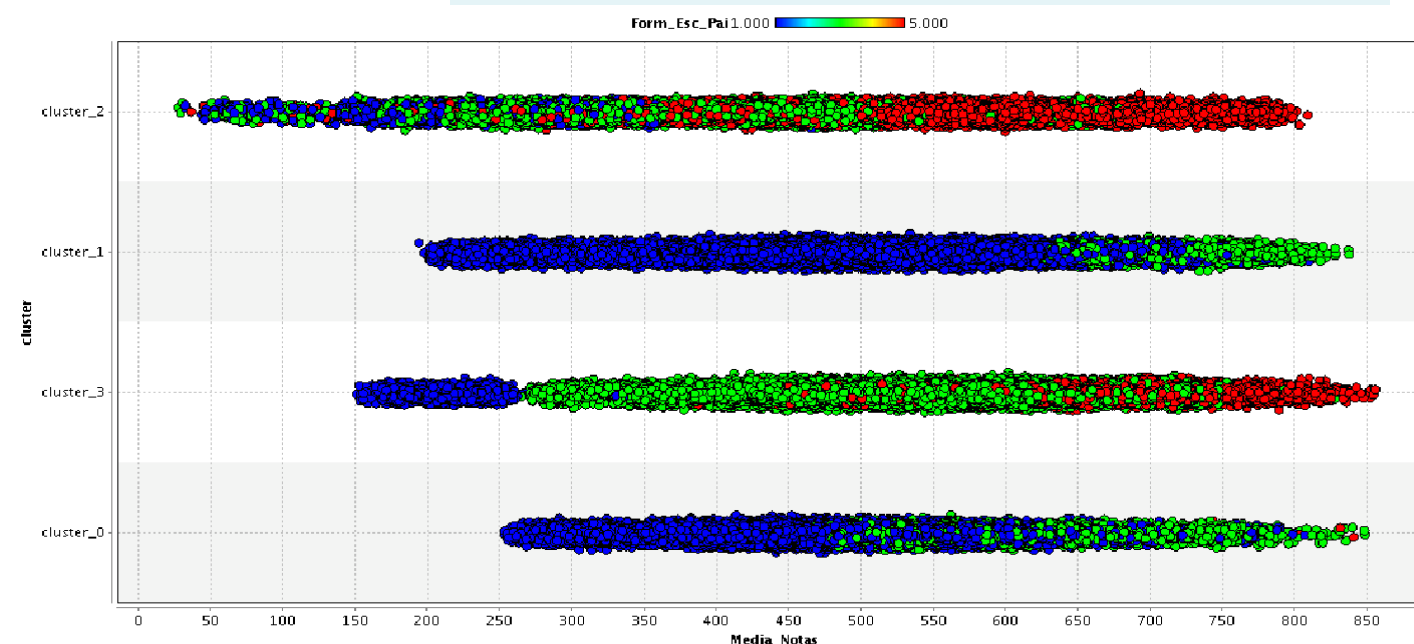
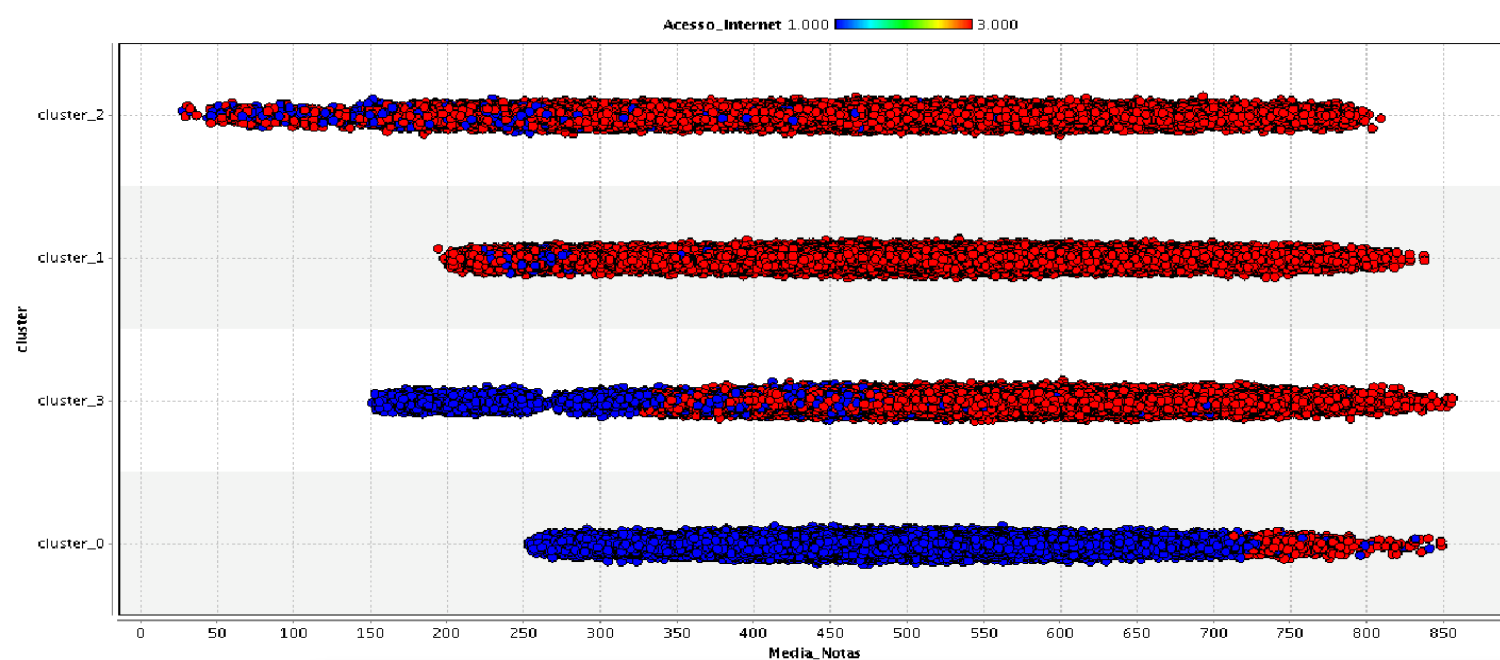
Usamos a tecnologia de mineração de dados de agrupamento para descobrir o perfil dos estudantes que fazem o exame e verificar se podem ou não impactar positivamente nas notas dos mesmos.

## OBJETIVOS

- Descobrir padrões ocultos, utilizando algoritmo de agrupamento k-means, aplicado ao conjunto de dados ENEM 2018;
- Implementação de código para as fases de limpeza, transformação e codificação dos dados;

## MATERIAIS E MÉTODOS

Arquivo disponibilizado pelo INEP com os microdados do ENEM 2018;  
 Python para limpeza, transformação e codificação dos dados;  
 Orange para geração de amostra dos dados;  
 Finalmente o RapidMiner para aplicação do algoritmo de agrupamento;



Cores			
Eixo horizontal secundário	1	3	5
Renda Familiar	Classe Baixa	Classe Média	Classe Alta
Formação Escola Pai	Ensino Fundamental	Ensino Médio	Ensino Superior

Cores			
Eixo horizontal secundário	1	Não se aplica	3
Acesso a Internet	Não possui Internet	Não se aplica	Possui Internet

## RESULTADOS

- Descobrimos que apesar de não ser condição essencial, aspectos como escolaridade dos pais, se possui acesso à internet e renda da família, tem indícios de afetarem a média do estudante.