

Competição de Ciência de Dados do Departamento de Estatística

Equipe: RPy; **Integrantes:** Wellerson Oliveira, Ianca Leite e Matheus Pessoa

Análise exploratória do Censo Educacional 2017

1- INTRODUÇÃO

A competição disponibilizou o censo educacional de 2017 para realização das análises e extração de dados. Partindo desse *dataset* e da exploração de suas colunas e valores, foram decididos quais dados seriam analisados, tendo em vista que o conjunto disponível era imenso, cerca de dezesseis *gigabytes*. Para possibilitar um contexto melhor para os dados, foram usados como base também as notas das avaliações de desempenho do ensino médio e fundamental, SAEB e ENEM, obtidas da plataforma INEP. Apenas as escolas com notas cadastradas foram selecionadas.

2- ANÁLISE DA INFRAESTRUTURA DAS ESCOLAS

Nessa análise utilizamos os dados de infraestrutura das escolas analisadas pelo censo e cruzamos esses dados com as notas de duas provas: SAEB e ENEM. O objetivo foi tentar identificar quais características mais influenciavam nas notas dessas provas. Para isso, selecionamos as variáveis que julgamos ser mais importantes para a análise e buscamos um algoritmo de aprendizagem de máquina para criar um modelo e medir a influência dessas variáveis no modelo.

O algoritmo escolhido foi o *random forest* pois ele oferece uma boa precisão com dados de entrada categóricos. As escolas tinham informações de infraestrutura como a existência de conexão à internet, biblioteca ou sala de leitura, se a escola provê alimentação, entre outros. Executamos quatro vezes o algoritmo, no total: Uma para o modelo da prova SAEB para o fundamental 1, outra para o fundamental 2, outra para a nota do ENEM das escolas públicas e por fim uma para as escolas privadas.

3 - ANÁLISE DE APTIDÃO DOS PROFESSORES DO ENSINO MÉDIO

Além da análise da infraestrutura das escolas, foi feita também uma checagem acerca da formação dos professores em relação à disciplina ministrada por eles. O objetivo dessa análise foi identificar a porcentagem de professores no ensino médio que ministram disciplinas incompatíveis com sua formação. Fizemos então uma análise mais simplista, onde consideramos aptos apenas os professores com formação na área da disciplina ministrada (ex.: Professores de química com formação em Licenciatura em Química ou Bacharelado em Química). Fizemos essa análise separando escolas públicas e privadas. A partir disso, pode-se perceber que disciplinas técnicas como química e física têm, em geral, menos professores aptos em exercício.