

UNIVERSIDADE FEDERAL DE GOIÁS
FACULDADE DE INFORMAÇÃO E COMUNICAÇÃO
CURSO DE GESTÃO DA INFORMAÇÃO

Disciplina de Mineração de Dados - 2021-2

Prof. Dr. Alex Fabianne de Paulo

TRABALHO PRÁTICO

Esta atividade final da disciplina de mineração de dados traz um estudo de caso baseado em dados abertos sobre a educação básica no Brasil disponibilizados pelo Ministério da Educação. O trabalho consiste em realizar todo o processo de descoberta do conhecimento baseado na metodologia CRISP-DM. Lembrem-se de percorrer todas as etapas da metodologia:

1. Entendimento do negócio
2. Compreensão dos dados
3. Preparação dos dados (seleção e tratamento das variáveis selecionadas)
4. Mineração (modelagem baseada na tarefa de associação)
5. Interpretação e avaliação dos resultados.

O problema que se tem é o seguinte: existem associações entre variáveis que permitem identificar um padrão predominante?

Assim, o objetivo final é analisar e caracterizar as principais regras que associam as faixas de IDEGES e com demais variáveis. Deve ser definido um conjunto de regras associativas que demonstram padrões de combinação das variáveis, caso exista. É preciso justificar o porquê das escolhas de tais regras ou porque não foram identificados padrões.

ATENÇÃO: a única variável obrigatória para estar no modelo é a Faixa_IDEGES. A busca por padrões deverá se basear na combinação dos valores desta variável com as demais variáveis disponíveis na base de dados.

As duplas precisam mostrar cada etapa da metodologia, bem como os resultados obtidos que caracterizem as regras de associação. A definição das melhores regras de associação faz parte da atividade. O mesmo para a definição das variáveis que serão utilizadas. Análises exploratórias, como por exemplo por meio de *bloxplot*, histogramas, análise de frequência, etc., poderão ser usadas para complementar os resultados.

Juntamente no arquivo com a base de dados, é fornecido um dicionário de dados com o significado das variáveis. Leiam com atenção o dicionário para compreender o real significado das variáveis.

A **apresentação** dos resultados ocorrerá na aula do **dia 05 de abril de 2022**, com duração de **20 minutos por dupla**. Deverá ser entregue neste mesmo dia via tarefa no SIGAA, os códigos em Python (ou outras ferramentas utilizadas), bem como a apresentação com os resultados e análises desenvolvidos.

Atenção:

- i. Será considerado na nota a apresentação oral dos resultados obtidos, a coerência das análises, a qualidade e organização do material, o cumprimento do prazo e a entrega via SIGAA.

- ii. Dado a diversidade de variáveis, espera-se resultados e interpretações diferentes entre os times. Assim, resultados similares poderão ser desconsiderados e os alunos envolvidos terão as notas desta atividade zeradas.