

FACULDADE DOM BOSCO DE PORTO ALEGRE Credenciada pela Portaria Nº 3.254, de 26 de Novembro de 2002 Curso de Sistemas de Informação (Reconhecido pela Portaria 1093 de 24/12/2015, publicado no DOU de 30/12/2015)	
Disciplina: Introdução a Ciência de Dados	
Valor: 30%	Data: 01/12/2022
Professor: Adriana Paula Zamin Scherer	
Projeto Final da Disciplina – Parte 1	

Orientações:

1. Este trabalho deve ser entregue **até as 19h:10min do dia 09/12/2022** e tem peso 3.0. Não serão aceitos trabalhos entregues fora do prazo, pois essa é a última aula do semestre.
2. A entrega deverá ser feita através do link da tarefa no Moodle da disciplina.
3. O nome do arquivo deve ser ProjetoParte1<nome do aluno>.doc.

Tarefa 1 - Definição do escopo do trabalho:

- 1) Escolha uma base de dados que será utilizada no trabalho. Algumas **sugestões** de sites que possuem bases de dados para utilização com algoritmos de *machine learning*:
 - a) <https://archive.ics.uci.edu/ml/index.php>
 - b) <https://www.kaggle.com/>
 - c) <https://dados.gov.br/>
- 2) Alternativamente, serão aceitos os *datasets* do pacote *datasets* do R, neste caso, deverá ser um para cada tipo de aprendizado trabalhado na disciplina.
- 3) Explique de forma geral do que se trata(m) a(s) base(s) escolhida(s);
- 4) Crie o dicionário de dados para sua(s) base(s) de dados, explicando o significado de cada uma das variáveis que compõem a(s) base(s) de dados.
- 5) Defina o problema central.

Nota: A base teórica estabelece a definição do problema como etapa inicial de um processo de ciência de dados. No entanto, por se tratar de um trabalho acadêmico e a proposta deste trabalho é explorar *datasets* prontos, o início de deu a partir da escolha do *dataset*.

- 6) Explique o tipo de aprendizado e o algoritmo inicial utilizado para a possível solução do problema a ser resolvido.