

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 2	

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 a ser entregue deve ser ProjetoParte2<nome do aluno>.zip, que deverá conter: (a) o arquivo ProjetoParte2<nome do aluno>.docx com a documentação do projeto; (b) o arquivo ProjetoParte2<nome do aluno>.R com os comandos R para a implementação do projeto; (c) o arquivo que contém a base de dados utilizada.

Tarefa 2 – Implementação do projeto e documentação:

- 1) Desenvolver análises que contribuam para o conhecimento da base estudada. Nessa tarefa devem ser aplicados os conhecimentos relativos à análise exploratória com uso da estatística descritiva (material da Aula3 dará suporte a esta atividade, bem como o segundo trabalho avaliativo).
 - a. Requisito mínimo: 2 análises quantitativas e 2 análises visuais.
 - b. Todas as análises devem ser explicadas: O quê? Por quê? Como? Resultados.
- 2) Com base nas atividades desenvolvidas na Tarefa 1 - Definição do escopo do trabalho, implementar o script que suporte a solução do problema definido. Para tanto:
 - a. O script deve ser implementado utilizando a Linguagem R e deve conter comentários que auxiliem a compreensão da programação desenvolvida;
 - b. Devem ser seguidas todas as etapas inerentes ao tipo de aprendizagem e ao algoritmo escolhido para o desenvolvimento, conforme definições estabelecidas no estudo dos algoritmos de Machine Learning.
 - c. Caso você escolha a aprendizagem supervisionada, deverá ser realizada a avaliação de algoritmos através da comparação dos resultados de acurácia entre, pelo menos, dois

algoritmos diferentes (por exemplo: naive-bayes e árvore de decisão OU regressão linear e regressão com árvore de decisão OU demais combinações).

- d. Para aprendizagem não-supervisionada, a comparação não se faz necessária, visto que estudamos apenas um algoritmo para cada método de aprendizagem não-supervisionada.
- e. Para qualquer tipo de aprendizagem (tanto de aprendizagem supervisionada quanto a não-supervisionada), se faz necessário a execução de pelo menos 2 vezes o script com parâmetros diferentes, a escolha do cientista de dados e conforme os algoritmos permitam.
- f. Todos os experimentos devem ser explicados: O que foi feito? Por quê? Como? Que resultados foram alcançados? Qual o significado?

3) Conclusão final: relatar de forma argumentativa a análise dos resultados encontrados.