**Projeto de Avaliação**

**Especialização em Ciência de Dados**

**Big Data, Machine Learning e Aplicações**

**Módulo II - Disciplina: Machine Learning I**

**Overview\_\_\_\_**

Neste projeto, cada aluno deverá treinar, avaliar e interpretar um classificador.

**Detalhamento\_\_\_\_**

Cada aluno deverá escolher uma das 2 bases de dados disponibilizadas ([CancerMama.csv ou contratos-paraiba.csv](https://github.com/talitalobo/EspecializacaoUnifacisa2019/tree/master/dados)) e produzir um relatório que contenha:

* O contexto do que deseja modelar: qual sua pergunta alvo? Uma breve introdução sobre o problema abordado e como ele está posto nos tipos de aprendizagem vistos em sala.
* O processo de treinamento, avaliação e escolha do(s) modelos.
* Ao longo do relatório, descreva a metodologia empregada durante os experimentos: % de treino e teste, interpretação da matriz de confusão, métricas de avaliação escolhidas e suas aplicações no contexto do problema, interpretação dos resultados do modelo e suas predições, importância das features e os demais tópicos abordados em sala.
* Ao fim do relatório, descreva também suas impressões sobre o processo: quais suas maiores dificuldades? Quais os aprendizados?

*Atente que: mais importante do que o resultado obtido é a interpretação que você faz dele. Faça reflexões/discussões em relação ao contexto onde o problema está inserido, em como aquele resultado resolve ou não o problema que você imaginou.*

**Bônus:**

Utilizar uma técnica de agrupamento, interpretar os grupos gerados a partir do contexto dos dados. Descrever também a motivação para escolha do algoritmo.

**Envio\_\_\_\_**

* Data de Entrega: 22 de dezembro.
* O relatório **deve** ser entregue num Jupyter Notebook: seja o próprio relatório ou link público para o Github que o contém (relatórios inacessíveis não podem ser corrigidos).
* Email para envio: [vera.medeiros@maisunifacisa.com.br](mailto:vera.medeiros@maisunifacisa.com.br) .

**Considerações\_\_\_\_**

* O projeto é **individual**.
* Caso o aluno não envie o projeto, isso implicará sua reprovação no módulo.
* Identificado plágio, ambos os projetos serão zerados.
* Todo e qualquer código ou decisão metodológica tomada durante o desenvolvimento do relatório deverão ser explicados com base nos assuntos vistos em sala ou da bibliografia (caso necessário inserir fontes/referências).