

Introdução à Ciência dos Dados – Prova Prática 02

Professor: Fabrício A. Silva

Valor: 20 pontos

Data: 26/11/2020 (Entrega via PVANet)

Observações: é permitido usar os arquivos e códigos já desenvolvidos ao longo da disciplina, e consultar qualquer conteúdo na Internet. Não é permitida nenhuma interação com colegas.

Forma de entrega: Entregar um zip com os arquivos em formato .ipynb e html (ou PDF) contendo um roteiro com explicações e decisões que foram tomadas, e as visualizações e resultados gerados.

Leia o dataset “*Dados-Prova02.csv*” (PVANet), que contém informações de contratos de planos de assinatura de uma empresa de telecomunicações. Os nomes das colunas são auto-explicativos, contendo informações sobre o cliente (id-usuario, sexo, is-idoso, has-parceiro, has-dependentes), o contrato (tempo-contrato-meses, tipo-contrato, metodo-pagamento, valor-mensal) e os serviços que são assinados (has-seguranca-online, has-servico-telefone, has-multiplas-linhas, servico-internet, etc). Um atributo relevante, que será utilizado na questão 3, é o “Churn”, que indica se o usuário cancelou o seu contrato (valor=1) ou não (valor=0).

1. (5 pontos) Faça uma análise exploratória dos dados, para conhecer melhor sobre os mesmos, seus tipos, suas características, suas correlações, dentre outras informações que julgar necessárias e importantes.
2. (5 pontos) Utilizando regressão linear, faça um modelo para mostrar quais variáveis explicam melhor o valor do custo mensal (valor-mensal) que um usuário paga, e tentar prever esse valor com base em outros atributos. Avalie a qualidade do modelo, buscando encontrar as melhores combinações de atributos para essa previsão.
3. (5 pontos) Utilizando aprendizado supervisionado, crie modelos para prever se um usuário é potencial candidato a cancelar o contrato (Churn=1) ou não (Churn=0). Teste diferentes possibilidades de algoritmos e atributos, e avalie diferentes métricas.
4. (5 pontos) Utilizando aprendizado não-supervisionado, crie um modelo para separar os usuários em dois grupos (sem utilizar o atributo Churn). O objetivo é verificar se é possível utilizar esses agrupamentos para separar bem os usuários que irão cancelar o contrato ou não. Verifique a qualidade do agrupamento, e se foi possível ver semelhanças entre os grupos criados e os grupos reais de Churn=1 e Churn=0.