

Trabalho de aplicação - MAE0501

Nome do grupo: Grupo 2

Membros do grupo:

1. Alberto Rodrigues Ferreira
2. Willian Miranda A. da Silva

Motivação:

A principal motivação do grupo para realizar este trabalho é pelo fato de ser uma aplicação bem interessante em um contexto de classificação e pela grau de complexidade dos dados. Há várias abordagens de pré-processamento e modelagem para tentar obter bons resultados nesse tipo de problema. Pelo fato de se ter um número relativamente alto de observações nos conjuntos de treino/teste e um grande número de variáveis nos leva a acreditar que é um problema desafiador e bem atual.

Descrição do problema:

O problema desse desafio se resume a identificar clientes insatisfeitos do banco Santander no início de seu relacionamento. Isso permitiria que o banco tome medidas proativas para melhorar a felicidade do cliente e que previna a saída desses clientes insatisfeitos.

Descrição do conjunto de dados:

Como é uma competição do kaggle, existem dois conjuntos de dados: conjunto de dados de treino, em que faremos nossas análises e validações para ajuste dos modelos, que possui 76020 observações e 371 colunas anônimas, ou seja, nesta competição não é fornecido o significado das variáveis e também há o conjunto de teste, em que aplicaremos os mesmos pré-processamentos do conjunto de treino mas será usado para predição da variável resposta da competição, este conjunto possui 75818 observações e 370 variáveis, não possui a variável resposta justamente porque será usado para avaliação da performance do nosso melhor modelo.

Avaliação do problema:

É um problema de classificação, em que o objeto é prever se um cliente está satisfeito ou insatisfeito com o banco, baseado em sua experiência bancária. Alguns dos objetivos adicionais que não estão na competição é tentar entender o comportamento de clientes ficarem satisfeitos ou insatisfeitos.

Endereço da competição:

<https://www.kaggle.com/c/santander-customer-satisfaction/overview>