



O entendimento do problema de negócio

☰ Módulo	Ciclo 01: O problema de negócio e planejamento da solução
# Aula	4
☑ Ready	☑
☑ Finished	☑

Objetivo da Aula:

- ☐ Entender o problema de negócio.
- ☐ Apresentação do Conjunto de dados
- ☐ Próxima aula.

Conteúdo:

▼ 1. Problema negócio

A empresa de Seguro Insurance All oferece o serviço de Seguro Saúde para seus assegurados. O cliente paga um valor anual pelo seguro. O time de produto quer lançar um novo produto: o seguro de carro. E então gostariam de saber se o produto é viável. Para isso, eles entrevistam sua carteira de clientes perguntando se eles estariam interessados num seguro de automóvel. Vamos treinar um classificador para saber se as pessoas comprariam ou não esse seguro.

Mas, existe um limite de pessoas que serão contactadas via ligação telefônica e para as quais será oferecido o seguro de automóvel. Porém, o time de operações tem uma capacidade de fazer 20.000 ligações. Daí, surgem as perguntas

Quem são essas 20.000 pessoas para quem vamos ligar?
Como priorizar essa lista para que sejam feitas as ligações?

▼ 2. Conjunto de dados

<https://www.kaggle.com/datasets/anmolkumar/health-insurance-cross-sell-prediction>

- Crie uma pasta para o curso e guarde os dados numa subpasta chamada

`datasets`

▼ 3. Planejamento da solução com método SAPE

▼ SAída

- Aplicativo (via link na web) que recebe os dados de clientes novos e retorna a probabilidade de comprar o novo seguro

▼ Processo

- Baixar os dados e ler os dados
- Executar o processo de limpeza de dados como:
 - renomear colunas
 - checagem de valores NA
 - imputação de valores NA
 - renomear categorias
 - mudar de tipo de variável, caso necessário

- Estatística descritiva com visualização
- Criação de hipóteses e validação das mesmas
- Análise exploratória
 - Univariada
 - Bivariada
- Preparação dos dados para treinamento de modelos de Machine Learning (feature engineering)
- Seleção das variáveis mais importantes
- Treinamento de algoritmos de ML
- Tunagem de hiperparâmetros e cross validation
- Comparação da performance dos modelos
- Escolha do modelo final
- Uso do modelo final para fazer previsões de novos dados em um aplicativo na web

▼ **Entrada**

- Tabela de dados de clientes em .csv fornecida pela empresa