



# Semantix

AI for all



Thiago Vieira

Projeto de Parceria - Semantix

Curso - Profissão: Cientista de Dados

Este documento contém informações complementares para o arquivo python notebook contido neste mesmo notebook, onde serão mostradas todas as etapas e o código correspondente.

Este trabalho consiste em aplicar as etapas da metodologia Crisp-DM em uma base de dados, sendo essa metodologia dividida nas seguintes etapas:

- Entendimento do negócio
- Entendimento da base de dados
- Preparação da base de dados
- Modelagem
- Avaliação do Modelo
- Implementação

## **Objetivo**

Temos uma base de dados que trata sobre informações de clientes de uma empresa que oferece seguros de saúde. Essa empresa está estudando a possibilidade de oferecer também seguros automotivos, mas para isso, ela deseja estudar se haverá interesse por parte de sua clientela.

Para isso, construiremos um modelo de regressão logística e a partir deste iremos prever quais clientes de fato se interessam em adquirir tal seguro. Iremos também agrupar estes clientes em diferentes perfis para se ter uma ideia dos tipos de consumidor que mais se interessam na oferta, para ter uma ideia de uma possível campanha de marketing.

## **Coleta de dados, modelagem e conclusões**

As bases de dados utilizadas neste projeto são bases públicas encontradas no Kaggle, sendo que o link para elas encontra-se no notebook deste repositório.

Para a modelagem, como estamos lidando com uma variável binária, criamos um modelo de regressão logística na base de treino e usamos este para prever valores na base de testes.

Por fim, usamos a técnica de K Means para agrupar os clientes com base nos seus perfis como consumidores, além de estudar cada perfil.

Com isso, cumprimos o objetivo inicial do projeto de criar uma maneira para prever o interesse dos clientes na aquisição do seguro veicular.