



Machine Learning

Curso de especialização em Ciência de Dados e Big Data

Prof. Hugo de Paula

ATIVIDADE 2

PREPARAÇÃO DE DADOS

Adaptado de North, Matthew A. Data Mining for the Masses. 2012.

1. Carregue a base de dados Automoveis.xlsx, e realize a preparação dos dados para que eles sejam consistentes. Você irá encontrar:
 - a. Valores omissos que poderão ser registros ou atributos.
 - b. Valores omissos que poderão ser estimados.
 - c. Valores inconsistentes.
 - d. Problemas de formatação de dados.

Você poderá corrigir valores omissos e inconsistentes da forma que julgar procedente. Elimine campos caso ache que algum campo é redundante ou desinteressante. Explique sua estratégia. Quando é melhor substituir valores? Quando é melhor remover o registro? Quando podemos estimar o valor pelos demais?

FORMA DE ENTREGA

A atividade deve ser entregue na forma de um Notebook desenvolvido no Google Colab (<https://colab.research.google.com/>) e compartilhado com o e-mail hugodepaula@gmail.com, até 7 dias após a realização na aula.

Não é necessário fazer capa ou formatação no padrão ABNT.