

Projeto - Ciência dos dados - 2019.1

Para cada etapa abaixo, você vai precisar descrever detalhadamente suas ações e os resultados encontrados.

- 1- Escolha uma base de dados com pelos menos 10k elementos.
- 2- Faça o pré-processamento da base. Por exemplo, verificando se há informações faltantes, discretizando e normalizando os dados.
- 3- Gere as estatísticas descritivas de cada um dos atributos de dados e verifique se faz sentido os valores encontrados. A partir dessa análise, comece a gerar uma familiarização com os dados, intuindo o que pode ser verificado de correlação entre eles.
- 4- Gere informação visual de seus dados: construa pelo menos um gráfico de cada um dos 4 tipos de gráficos abaixo (use atributos diferentes para cada gráfico).
 - Para dados categóricos:
 - barras
 - Para dados numéricos:
 - histogramas
 - densidade
 - boxplot
- 5- Gere um conjunto de hipóteses (pelo menos 5) de seus dados. Por exemplo, se é possível fazer classificação ou previsão usando técnicas de IA. Outro exemplo: se uma dada discretização de um atributo numérico facilita a classificação dos dados.
- 6- Aplique pelo menos 3 técnicas de aprendizagem de máquina para analisar cada uma das suas hipóteses. Essas técnicas devem ser relevantes e atuais (Rede Neural, SVM, Árvore de decisão, Extreme Learning Machine, Redes Bayesianas, Redes profundas, etc). Confirme com o professor quais técnicas você irá utilizar.
- 7- Discuta sobre os resultados e as possíveis justificativas que levaram as técnicas a apresentarem seus resultados. Elas validaram sua hipótese?
- 8- Gere um relatório com a descrição das ações e resultados encontrados em todas as etapas do projeto.
- 9- Faça uma apresentação descrevendo suas ações no projeto.