

Avaliação

Questão 1

Use o conjunto de dados biogeoquímicos em (<https://www.sciencebase.gov/catalog/item/5fec8c1ed34ea5387defd59d>). Analise a normalidade de todos os dados numéricos, faça transformações quando necessário ou possível. Proponha um modelo usando os dados normais ou normalizados que tenha a concentração de sulfato como saída em função dos outros parâmetros.

Questão 2

Pesquise em bases de dados públicas informações para construir um classificador de áreas em risco de contaminação ou contaminadas em bacias hidrográficas. Seja criativo e procure por informações que tenham relevância, proximidade de grandes centros, fluxo, vazão, pluviometria, entre outros. Sugere-se usar dados de um mesmo país ou região. Tente aplicar alguma técnica de aprendizado de máquina.

Questão 3

Uma pequena digressão fora das geociências. Use os dados de propriedades de vinhos de diversas marcas que podem ser encontrados em (<https://archive.ics.uci.edu/ml/datasets/wine>). Tente estabelecer se é possível agrupar esses vinhos de alguma forma com essas propriedades.

Questão 4

Descreva em texto como você lida com um conjunto de dados que precisa analisar. Quais as principais etapas?