

Bootcamp IGTI: Analista de Machine Learning**Trabalho Prático****Módulo 3****Seleção de Modelos de Aprendizado de Máquina****Objetivos**

Exercitar os seguintes conceitos trabalhados no Módulo:

- ✓ Exercitar conceitos sobre medidas de desempenho para regressão..
- ✓ Modelar um problema como uma tarefa de regressão.
- ✓ Avaliar um modelo de regressão.
- ✓ Exercitar conceitos sobre medidas de desempenho para classificação.
- ✓ Modelar um problema como uma tarefa de classificação.
- ✓ Avaliar um modelo de classificação.
- ✓ Exercitar conceitos sobre medidas de desempenho para clusterização.
- ✓ Modelar um problema como uma tarefa de clusterização.
- ✓ Avaliar um modelo de clustering.

Enunciado

Neste trabalho, vamos exercitar conceitos sobre medidas de desempenho vistas em aula, a partir da modelagem de 3 problemas diferentes. Para o problema de regressão, usaremos a base **diabetes_numeric.csv** e uma **regressão linear**. Para o problema de classificação, usaremos a base **bloodtransf.csv** e um **SVM**. Para o problema de clusterização vamos utilizar a base **wine.csv** e o algoritmo **kmeans**.

Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Baixar os arquivos referentes as bases de dados e acessá-las pelo collab.
2. Obter informações sobre números de features e instâncias dos datasets.
3. Identificar a existência de dados faltantes nos datasets.
4. Separar os conjuntos de treino e teste, usando a função **train_test_slipt**, com **test_size = 0.25** e **random_state = 42**.
5. Importar o sklearn para:
6. Aplicar à base **diabetes_numeric.csv** o modelo de **regressão linear**.
7. Avaliar as métricas R2, MAE e MSE.
8. Aplicar à base **bloodtransf.csv** o modelo **SVC**, com **kernel=rbf**.
9. Avaliar as métricas Acurácia, Precision, Recall, F1 e AUROC
10. Aplicar à base **wine.csv** o modelo **kmeans**.
11. Identificar o número de clusters mais adequado de acordo com o dataset
12. Avaliar as métricas Coeficiente de Silhueta, Davies-Bouldin Score e Mutual Information.

Respostas Finais

Os alunos deverão desenvolver a prática e, depois, responder às seguintes questões objetivas: