

Bootcamp IGTI: Engenheiro(a) de Machine Learning

Desafio

Módulo 5	Desafio Final
-----------------	----------------------

Objetivos

Exercitar os seguintes conceitos trabalhados no Módulo:

- ✓ Pré-processamento dos dados.
- ✓ Detecção de anomalias.
- ✓ Processamento dos dados.
- ✓ Correlações.
- ✓ Redução da dimensionalidade.
- ✓ Algoritmos supervisionados e não supervisionados.

Enunciado

Para todo engenheiro de Machine Learning, é fundamental que todos os 7 passos para construção de uma solução sejam seguidos. Nesse sentido, saber percorrer todas essas etapas e, ao final, identificar a melhor estratégia é papel primordial para esse profissional.

Neste desafio final, vamos empregar boa parte dos conceitos mostrados no decorrer de todos os módulos do Bootcamp para a análise e a classificação de veículos do conhecido dataset “cars”. Esse dataset contém um conjunto de informações sobre vários veículos pesquisados. Existem dados, por exemplo, sobre a potência do veículo, sobre a origem e cilindradas cúbicas.

Para essa análise, vamos empregar os conceitos de redução da dimensionalidade com o PCA, clusterização com o K-Means e classificações com algoritmos supervisionados.

Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Acessar o link abaixo e realizar o *download* do arquivo “**cars.csv**”.

<https://drive.google.com/drive/folders/1nb7AcncQvt-OmGiTle7TepBaaZn30ZXN?usp=sharing>

2. Para a implementação dos algoritmos, utilizar as definições abaixo:

```
normaliza = StandardScaler() #objeto para a normalização
pca = PCA(n_components=7)
KMeans(n_clusters=3, random_state=42)
train_test_split(entradas_arvore, saidas_arvore,
test_size=0.30, random_state=42)
DecisionTreeClassifier(random_state=42)
LogisticRegression(random_state=42)
```

3. Para as questões que envolvem a construção de modelos supervisionados, você deve utilizar o dataset original para definir a eficiência dos veículos. Além disso, deve utilizar as variáveis ['cylinders' , 'cubicinches' , 'hp' , 'weightlbs' , 'time-to-60'] como entrada. A saída deve ser a classificação de eficiência do veículo.