

## CURSO SISTEMAS DE INFORMAÇÃO

### MINERAÇÃO DE DADOS

#### REGRESSÃO UTILIZANDO O PYCARET

Essa atividade tem por objetivo aplicar os conhecimentos de machine learning supervisionado a biblioteca PyCaret para resolver uma tarefa de **regressão** com um conjunto **de dados reais**. Dessa forma, é preciso que você explique todos os atributos utilizados do dataset.

Siga os seguintes passos:

- Importe e explore os dados escolhidos.
- Configure o ambiente do PyCaret para a tarefa de classificação, definindo a variável alvo.
- Compare os modelos disponíveis e selecione o melhor desempenho.
- Realize o ajuste de hiperparâmetros com `tune_model`.
- Avalie o modelo final, avalie as métricas mais importantes utilizando SHAP.
- Interprete os resultados: comente a performance do modelo, principais variáveis e possíveis limitações dos dados utilizados.
- Justifique a possibilidade desse modelo treinado ser utilizado ou não em um ambiente de produção.

#### Entrega:

Código Python comentado (.ipynb). Especifique dentro do código, a origem dos dados e a informação contida em cada atributo.

Na próxima aula, cada aluno deverá explicar o seu trabalho, inclusive todos os processos realizados pelo PyCaret.

Cada aluno apresentará o seu modelo treinado conforme a relação a seguir:

ALUNO	DATASET	PROBLEMA
ALFREDO DE SOUZA AGUIAR NETO	<a href="https://www.kaggle.com/datasets/aayushchou/hospital-length-of-stay-dataset-microsoft">https://www.kaggle.com/datasets/aayushchou/hospital-length-of-stay-dataset-microsoft</a>	Predição de tempo de internação hospitalar
ÁLVARO CALEBE SOUZA FERREIRA	<a href="https://www.kaggle.com/datasets/alamshihab075/mental-health-dataset">https://www.kaggle.com/datasets/alamshihab075/mental-health-dataset</a>	Predizer produtividade autoavaliada
ANA BEATRIZ MARQUES MOREIRA	Ana Beatriz Marques Moreira - <a href="https://www.openml.org/search?type=data&amp;status=active&amp;qualities.NumberOfClasses=lte_1&amp;id=46283">https://www.openml.org/search?type=data&amp;status=active&amp;qualities.NumberOfClasses=lte_1&amp;id=46283</a>	Prever o consumo de energia de eletrodomésticos numa residência
CAMILLE GOMES DE MENEZES (VICTOR HUGO)	<a href="https://www.kaggle.com/datasets/mohamedasak/imdb-top-250-movies">https://www.kaggle.com/datasets/mohamedasak/imdb-top-250-movies</a>	IMBD Top 250 Movies
CHARLES ABRAÃO PANTOJA	<a href="https://www.kaggle.com/datasets/devansodariya/student-performance-data">https://www.kaggle.com/datasets/devansodariya/student-performance-data</a>	prever a performance dos estudantes



SANTOS		
FELIPE FERREIRA A SOUSA	<a href="https://www.kaggle.com/code/leomauro/regress-o-pre-o-de-venda-da-casa">https://www.kaggle.com/code/leomauro/regress-o-pre-o-de-venda-da-casa</a>	Preço de venda da Casa
FILIPE DIAS BARRETO	<a href="https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities">https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities</a>	Prever a temperatura média diária de uma cidade
GABRIEL ALVES DA SILVA	<a href="https://www.kaggle.com/code/gustavolenin/preco-de-combustiveis-2004-a-2020">https://www.kaggle.com/code/gustavolenin/preco-de-combustiveis-2004-a-2020</a>	preço de combustíveis - 2004 a 2020
GUILHERME SILVEIRA FIGUEIRA	<a href="https://www.kaggle.com/datasets/toramky/automobile-dataset">https://www.kaggle.com/datasets/toramky/automobile-dataset</a>	Prever o valor de mercado de automóveis
HEMILLY CHRISTI NNE SILVA PEREIRA	<a href="https://www.kaggle.com/datasets/rudraprasadbhuyan/spotify-dataset-for-ml-practice">https://www.kaggle.com/datasets/rudraprasadbhuyan/spotify-dataset-for-ml-practice</a>	Spotify Dataset for ML Practice
HENRIQUE PONTES BARROS AMORIM	Henrique Pontes - <a href="https://www.kaggle.com/datasets/spittman1248/cdc-data-nutrition-physical-activity-obesity">https://www.kaggle.com/datasets/spittman1248/cdc-data-nutrition-physical-activity-obesity</a>	Compreender como hábitos alimentares
ISAC PEREIRA DA SILVA	<a href="https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009">https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009</a>	Prever o nível de qualidade de um vinho
JOANA		



LETÍCIA AZEVEDO DO NASCIME NTO SOARES	<a href="https://www.kaggle.com/datasets/lunthu/steam-monthly-average-players?resource=download">https://www.kaggle.com/datasets/lunthu/steam-monthly-average-players?resource=download</a>	
JOÃO VICTOR DE OLIVEIRA CARNEIR O	<a href="https://www.kaggle.com/datasets/safaruzzamanshovo/graduate-admission-dataset">https://www.kaggle.com/datasets/safaruzzamanshovo/graduate-admission-dataset</a>	Graduate Admission Dataset
JOÃO VICTOR FERNAN DES CORDEIR O MARTINS		
JOÃO VICTOR PÓVOA FRANÇA	<a href="https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017">https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017</a>	International football results from 1872 to 2025
JÚLIA PEREIRA GONÇAL VES	<a href="https://www.kaggle.com/datasets/wickedbotz/teste-desmatamento-amazonia">https://www.kaggle.com/datasets/wickedbotz/teste-desmatamento-amazonia</a>	Desmatamento da Amazônia
KAIO MACEDO MACHAD O		
KATRINA	<a href="https://www.kaggle.com/datasets/nikhil7280/student-performance-prever-o">https://www.kaggle.com/datasets/nikhil7280/student-performance-prever o</a>	



DE MORAIS SANTOS	multiple-linear-regression?utm_source=chatgpt.com	índice de desempenho de estudantes
LEANDR O GUIMARÃ ES GARCIA	<a href="https://www.kaggle.com/datasets/dongeorge/beer-consumption-sao-paulo">https://www.kaggle.com/datasets/dongeorge/beer-consumption-sao-paulo</a>	Beer Consumption
LUCAS ARAUJO DE CASTRO	<a href="https://www.kaggle.com/datasets/dhrumilgohel/student-performance-dataset">https://www.kaggle.com/datasets/dhrumilgohel/student-performance-dataset</a>	Prever a nota final do estudante
LUCAS DANIEL RODRIG UES DOS SANTOS	LUCAS DANIEL RODRIGUES DOS SANTOS  Dataset: NFL Stats 2012-2024 <a href="https://www.kaggle.com/datasets/philiphyde1/nfl-stats-1999-2022">https://www.kaggle.com/datasets/philiphyde1/nfl-stats-1999-2022</a>	Prever quais times jogariam a pós temporada da NFL.
LUCAS HENDER SON VIEIRA CUNHA	<a href="https://www.kaggle.com/datasets/denkuznetz/taxi-price-prediction">https://www.kaggle.com/datasets/denkuznetz/taxi-price-prediction</a>	Problema: Prever o preço de corridas de táxi com base em diversos fatores, como distância, horário, condições de tráfego e outros.



LUÍS EDUARD O NOLETO DE MELO		
MARCEL O ALVES LEITE	<a href="https://www.kaggle.com/datasets/rhythmcam/pycaret-regression-auto-model/data">https://www.kaggle.com/datasets/rhythmcam/pycaret-regression-auto-model/data</a>	Prever a quantidade de itens vendidos
MATEUS TOSHIO COLI SHIBATA		
OTAVIO NARDINI DE FARIA	<a href="https://www.kaggle.com/datasets/rehan497/worldwide-temperature-trends-19612024">https://www.kaggle.com/datasets/rehan497/worldwide-temperature-trends-19612024</a>	
VICTOR ALVES DOS REIS DIAS	Victor Alves Conjunto de dados Auto MPG <a href="https://www.openml.org/search?type=data&amp;status=active&amp;id=42372">https://www.openml.org/search?type=data&amp;status=active&amp;id=42372</a>	