

Inteligência Artificial - CST ADS

Professora Sandra Cristina Costa Prado

Exercício 2

04/10/2024

Conceitual:

1) Tratamento de valores ausentes e padronização dos preditores são técnicas de pré-processamento de dados. Por que estas técnicas são importantes? 0,5

2) Defina sobreajuste (overfitting) e sub-ajuste (underfitting) em modelos de aprendizado de máquina e diga um sinal indicativo de sobreajuste e sub-ajuste. 0,5

3) O que é validação cruzada? 0,5

Prática: Compreenda os dados, faça o pré-processamento, treine e avalie um modelo de aprendizado de máquina utilizando o dataset CO2Emission. Treine um modelo de Regressão Linear e um de Random Forest. O objetivo é prever a quantidade de gás carbônico que sai pelo escapamento. Passos: 1 - Importe os pacotes necessários, 2 - Carregue e exiba os dados, 3 - Compreenda e descreva os dados através de métodos do dataframe e histogramas, juntamente com as funções skew e kurtosis, 4 - Divida os dados em treinamento e teste, utilizando o parâmetro stratify da função train_test_split() para que os conjuntos de teste e de treinamento tenham a mesma proporção de tipos de combustível do dataset original, 5 - Faça gráficos de dispersão e utilize o método corr() do dataframe para encontrar correlações dos preditores entre si, e dos preditores com o target, 6 - Faça combinações de preditores, 7 - Separe os preditores e o target, 8 - Verifica se há dados ausentes e resolva se houver, 9 - Faça o tratamento de dados categóricos, 10 - Padronize os dados com StandardScaler, 11 - Treine o modelo e avalie utilizando a margem de erro (rmse) e 12 - Avalie com o conjunto de dados pré-processado, utilizando a margem de erro (rmse). 8,5