

Trabalho Final - Data Science e IA

Nome: Sílvia Rodolfo Melo de Sá Leitão

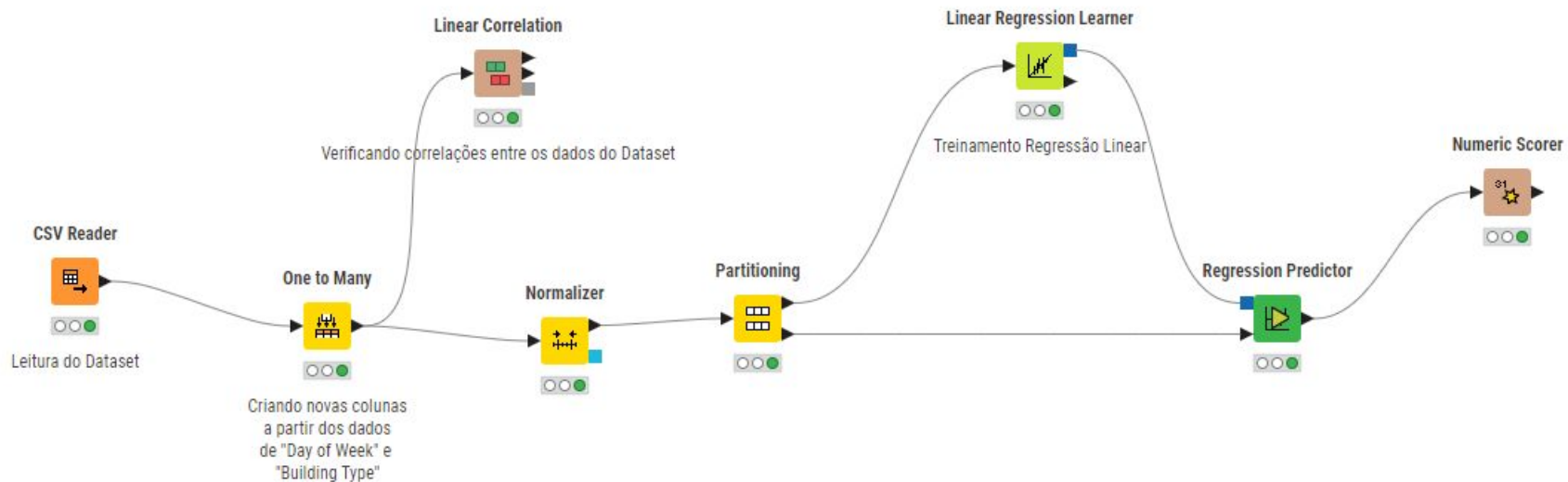
e-mail: srmsl@cesar.school

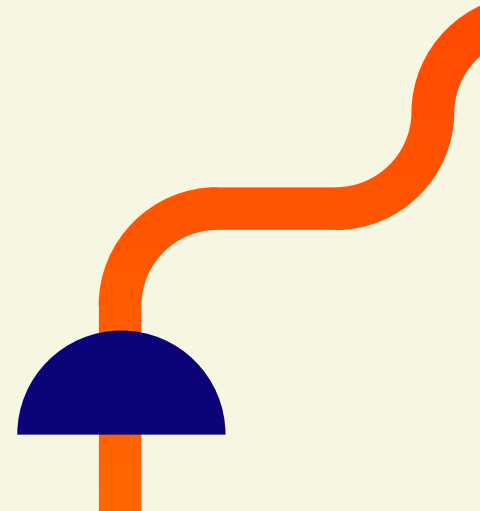
Link dataset:

<https://www.kaggle.com/datasets/govindaramsriram/energy-consumption-dataset-linear-regression>

RESOLUÇÃO DO DATASET VIA FERRAMENTA KNIME

Fluxo para o Treinamento do Modelo de Dados



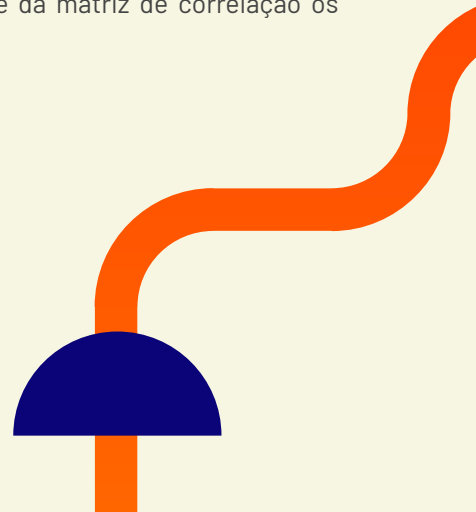
[illegible]

RESOLUÇÃO DO DATASET VIA FERRAMENTA KNIME

Este Dataset foi projetado para prever o consumo de energia com base nas características da Edificação e os fatores ambientais, O objetivo é construir um modelo preditivo para estimar o consumo de energia usando esses atributos. Dada a sua simplicidade do dataset o modelo de treinamento escolhido foi o de regressão linear e ferramenta utilizada para o treinamento deste dataset foi o KNIME.


O fluxo executado na ferramenta utilizou os seguintes nós:

- CSV Reader - Leitura do arquivo de dataset, para essa atividade foi utilizado o arquivo **"train_energy_data.csv"**
- One to Many - Transformação os dados das colunas "day of Week" e "Building Type" (ambas tipo **string**) em novas colunas do tipo **integer**
- Linear Correlation - Utilizado para identificar colunas altamente correlacionadas e remover possíveis redundâncias.
- Normalizer - Utilizado para normalizar os valores numéricos
- Partitioning - Utilizados para dividir os dados em treinamento e teste. O particionamento utilizado nessa atividade foi de
- Linear Regression learner - Utilizado para treinar o modelo com os dados de treinamento. Após a análise da matriz de correlação os seguintes campos do dataset foram utilizados:
 - Square Footage
 - Number of Occupants
 - Appliances Used
 - Average Temperature
 - Commercial (flag para edificações Comerciais)
 - Industrial (flag para edificações Industriais?)
 - Weekday
 - Weekend
- Regression Predictor - Utilizado para prever os dados de teste.
- Score - Análise das Métricas como acurácia, precisão, recall, e F1-score.





RESOLUÇÃO DO DATASET VIA FERRAMENTA KNIME

Após a finalização do treinamento os seguintes dados estatísticos foram obtidos no nó **SCORER**:

► 1: Statistics  Flow Variables

Rows: 7 | Columns: 1

Table  Statistics 

<input type="checkbox"/>	#	RowID	Prediction (Energy Consumption) <i>Number (double)</i>
<input type="checkbox"/>	1	R^2	1
<input type="checkbox"/>	2	mean absolute error	0.012
<input type="checkbox"/>	3	mean squared error	0
<input type="checkbox"/>	4	root mean squared error	0.014
<input type="checkbox"/>	5	mean signed difference	-0
<input type="checkbox"/>	6	mean absolute percentage error	0
<input type="checkbox"/>	7	adjusted R^2	1