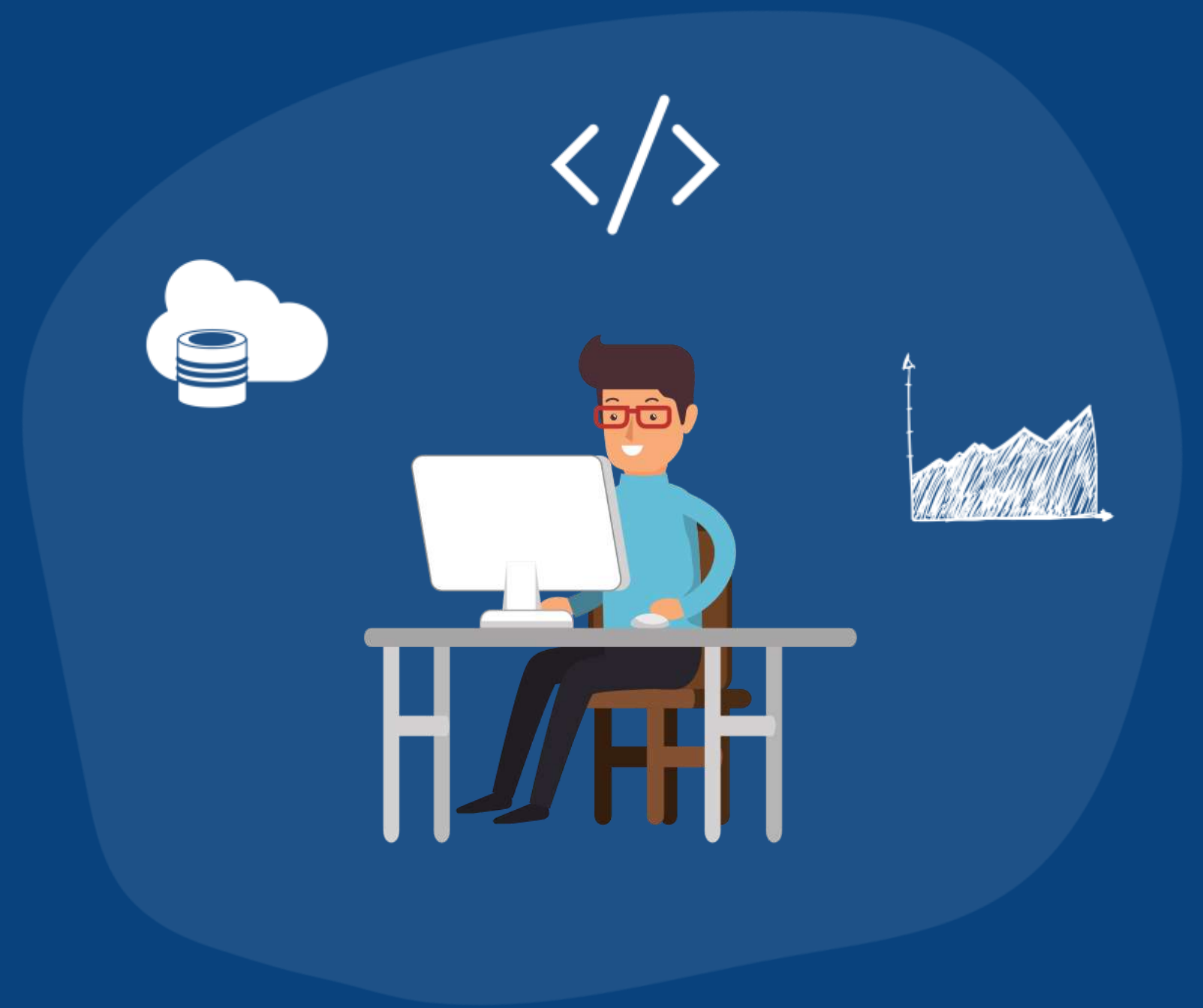


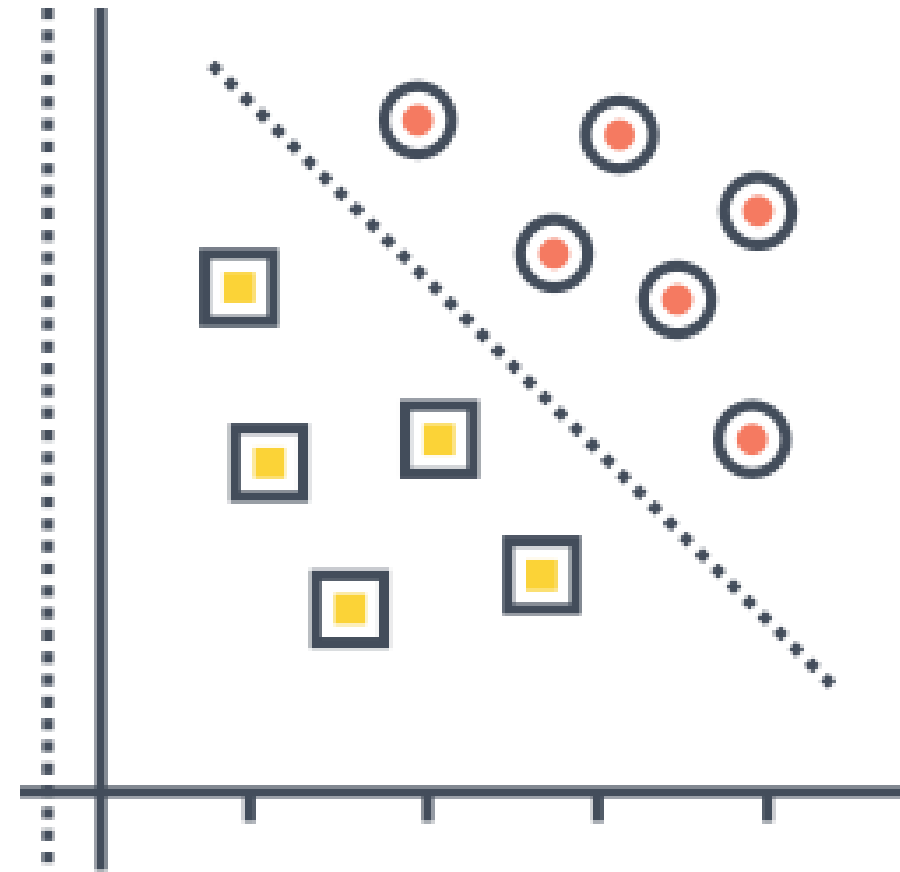
Módulo III:

Clusterização





Definição



CLUSTERING

É uma técnica de aprendizagem NÃO supervisionada em que dividimos pontos similares em grupos (ou clusters) onde a similaridade entre pontos do mesmo cluster é a maximada e a similaridade entre pontos de clusters distintos é minimizada.

Importância





CLIENTES

É uma ótima estratégia para entender como seus clientes podem ser agrupados, as peculiaridades e preferências de cada grupo, além da diferença entre eles. Isso permite rodar campanhas de marketing diferentes, oferecer produtos específicos, facilitar a construção de Personas, aumentar a receita da empresa e muito mais!



PRODUTOS

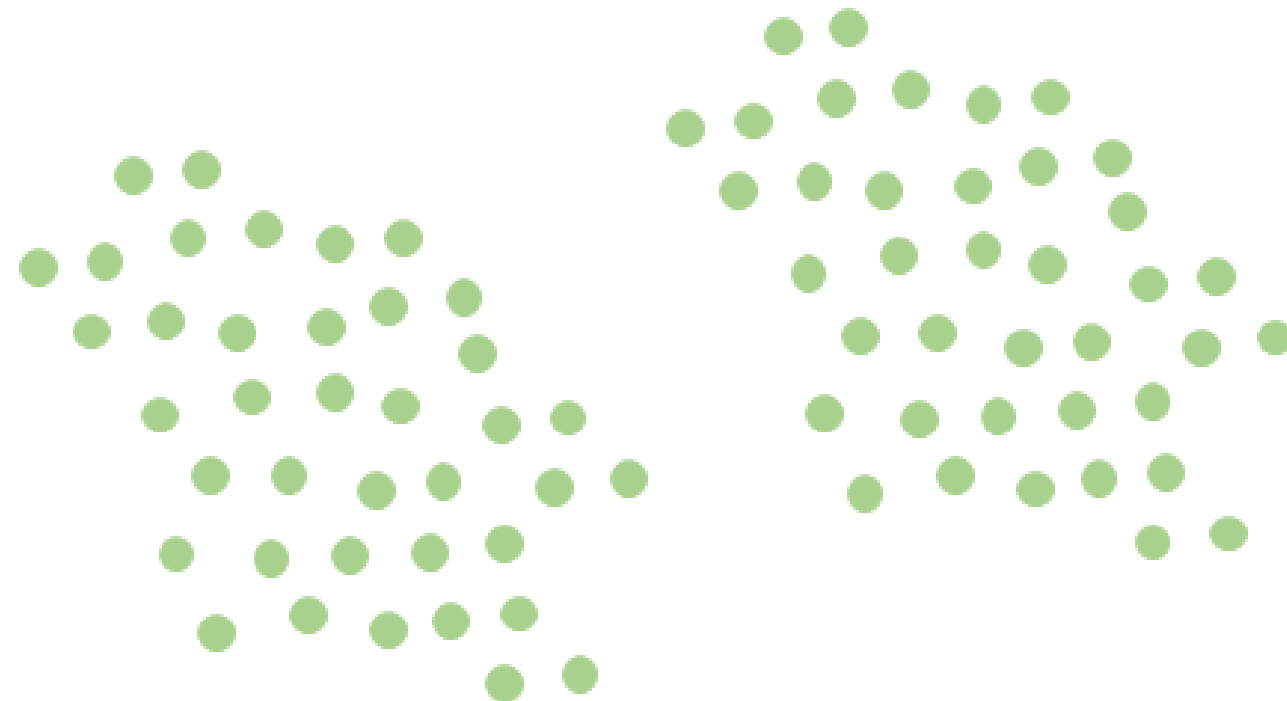
Clustering também pode ser usada para categorizar diversos produtos dentro de uma empresa ou no mercado. Às vezes temos diversas variáveis sobre esses produtos, mas não a categoria de cada um. Assim, o algoritmo pode trabalhar para encontrar os padrões e rotulá-los facilmente para você.

Algoritmo



INÍCIO

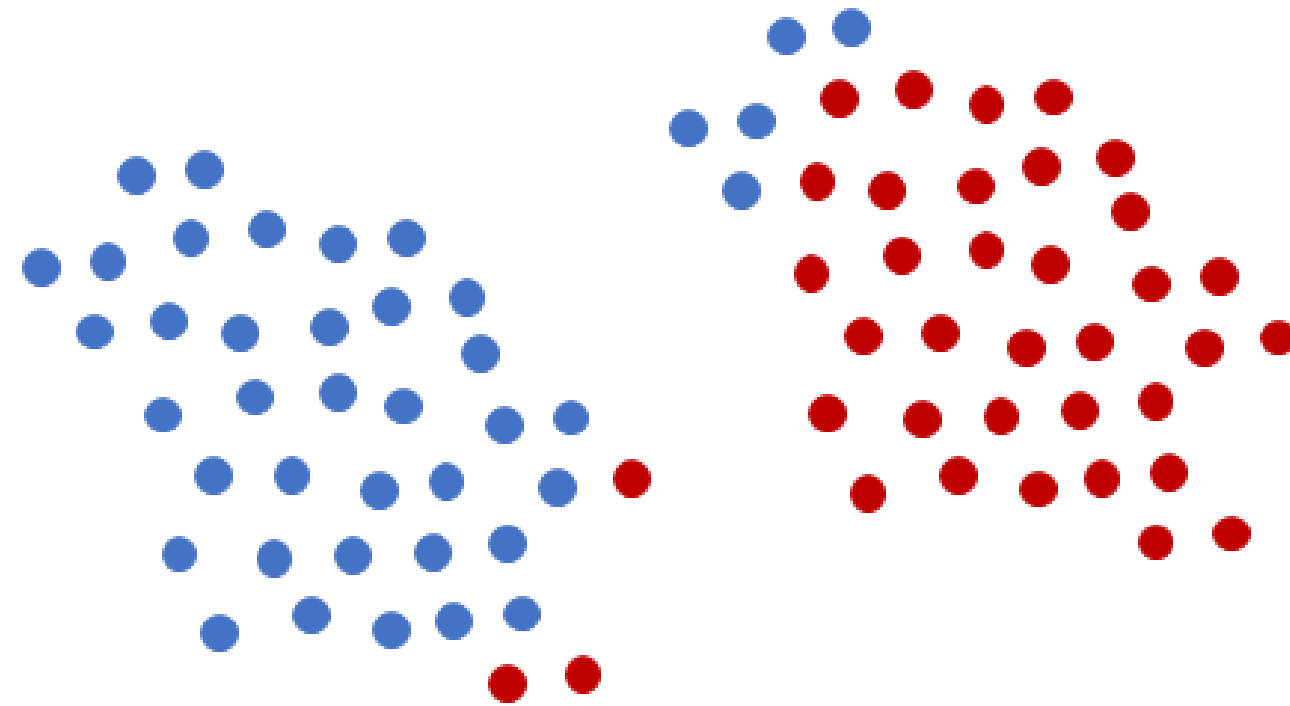
centróide 1



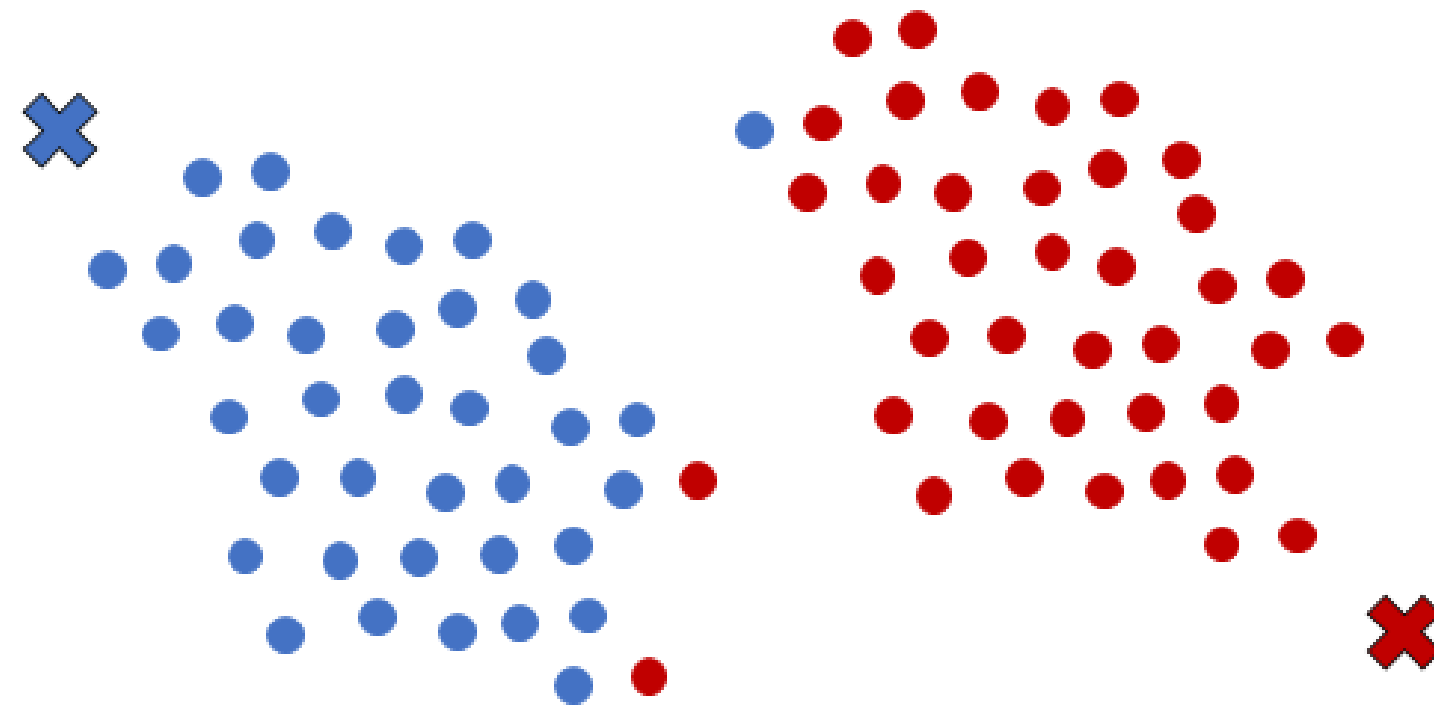
centróide 2



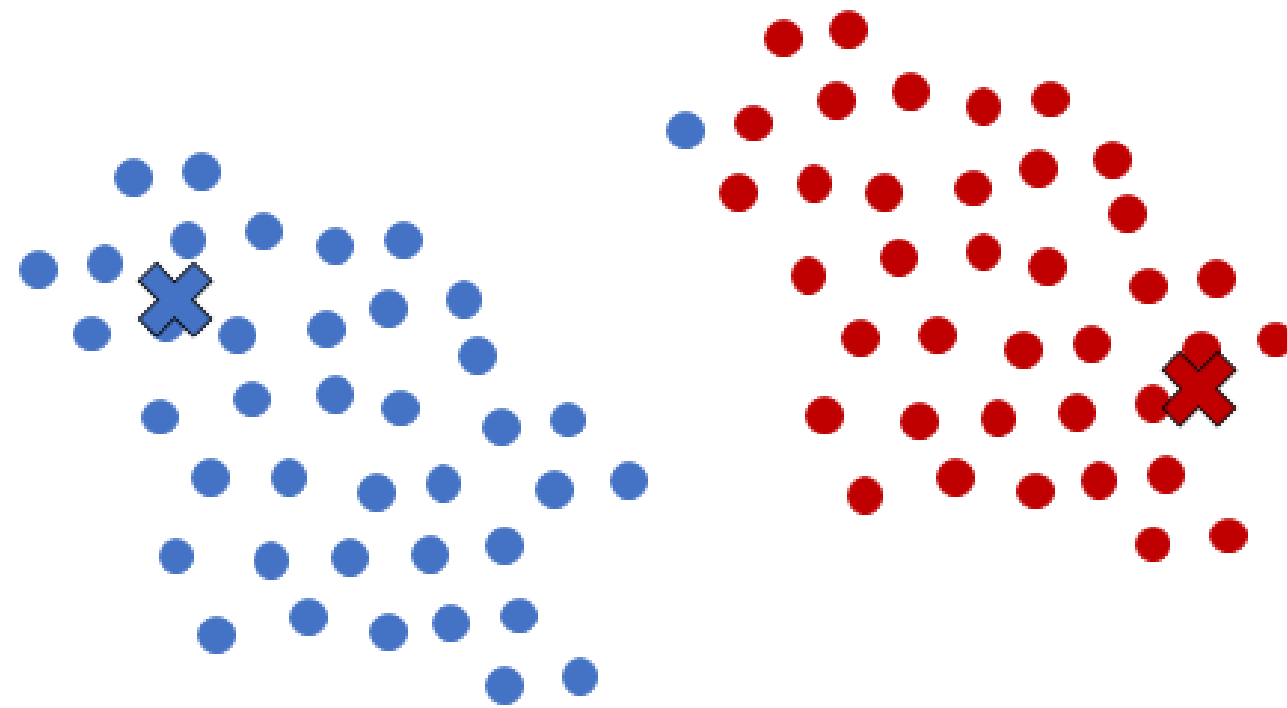
ITERAÇÃO 1



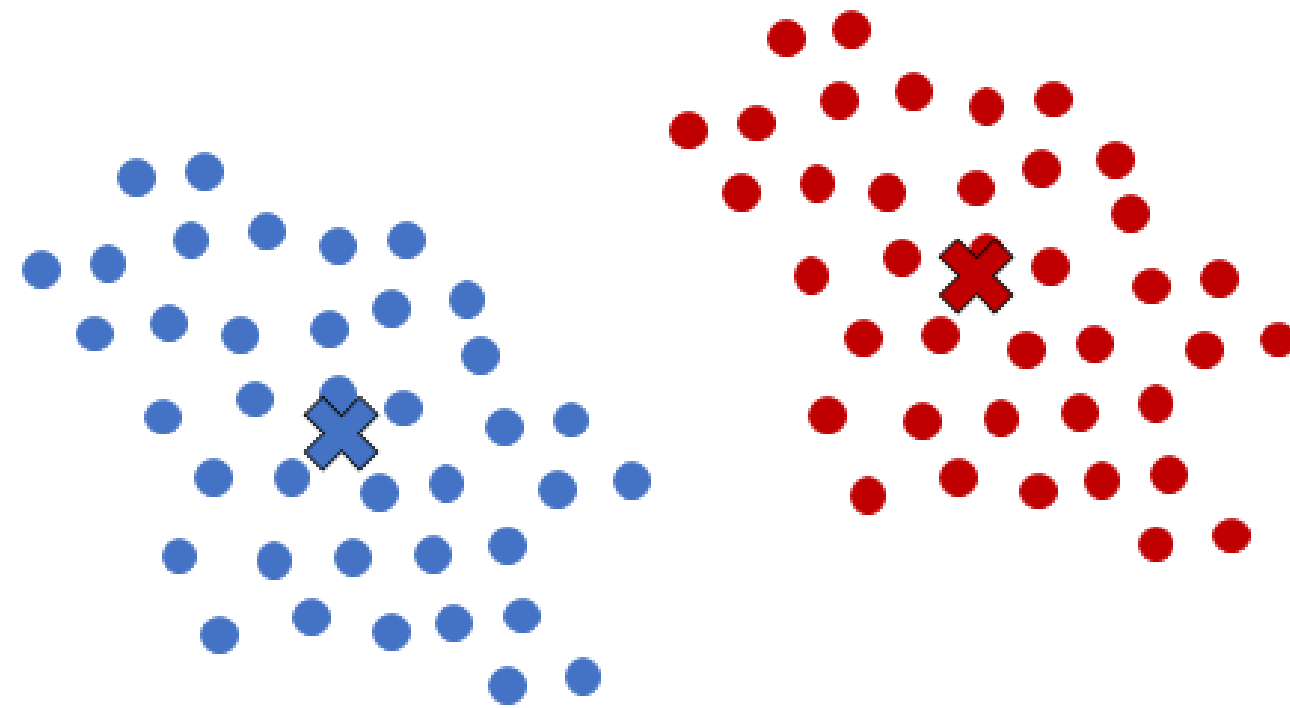
ITERAÇÃO 2



ITERAÇÃO 3

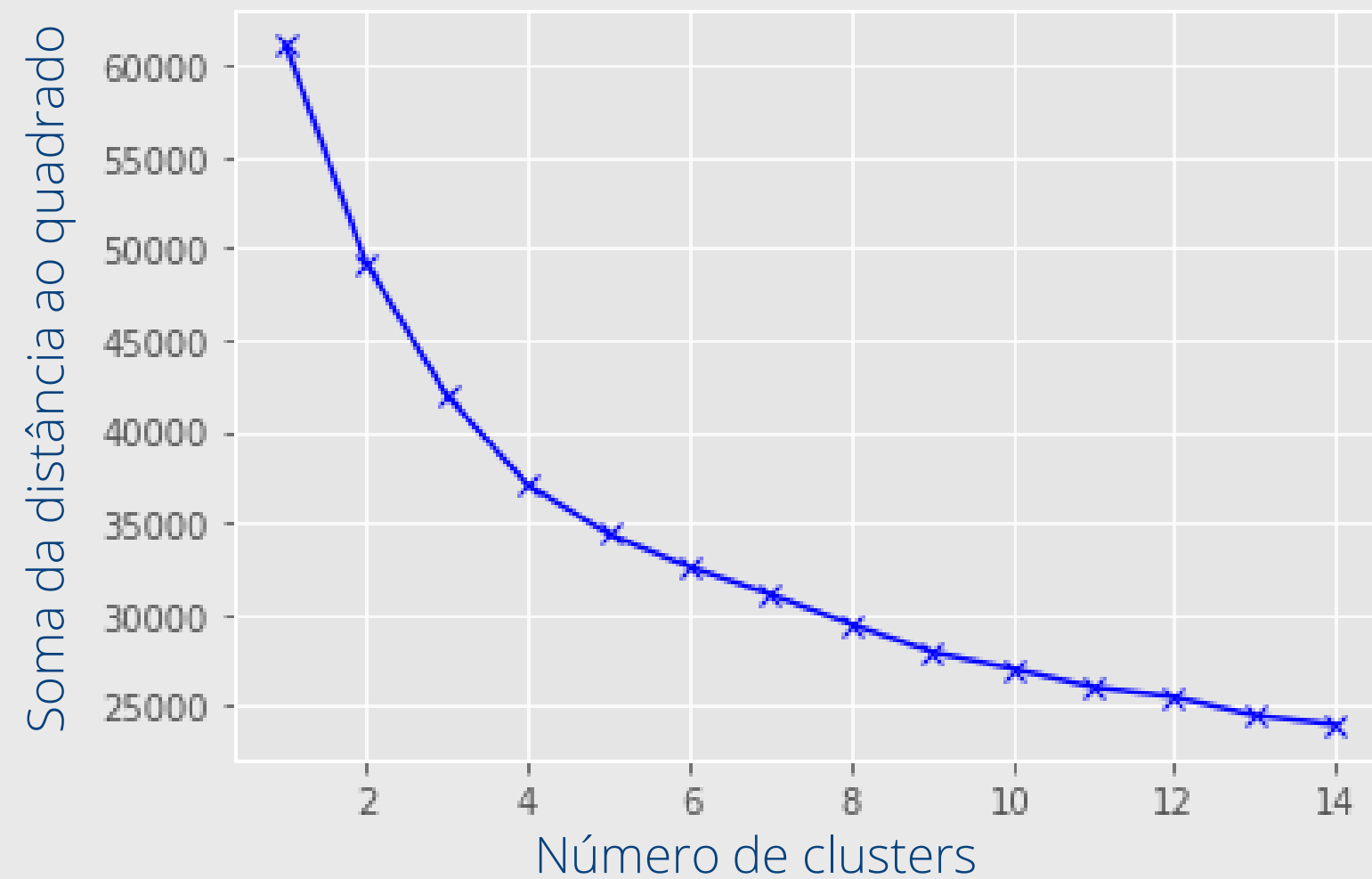
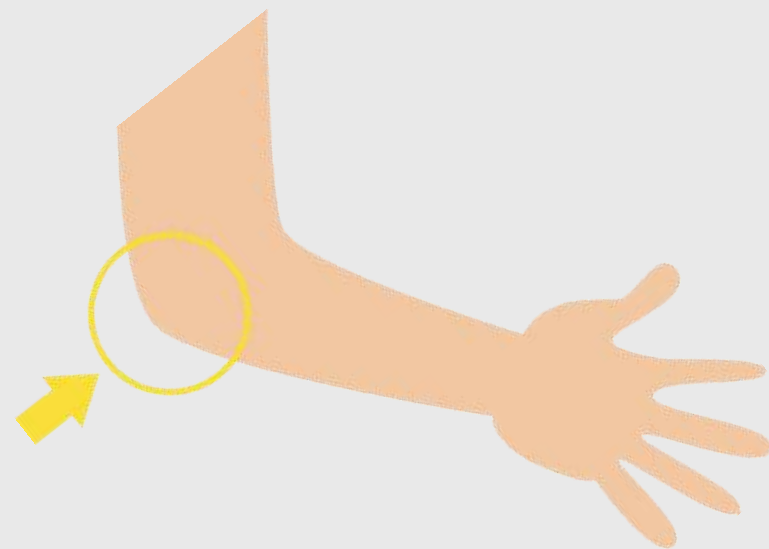


ITERAÇÃO 4 - FINAL





Método de Elbow



MÉTODO DE ELBOW

Esse método calcula a soma da distância ao quadrado entre cada ponto do cluster ao centróide, para cada valor de número de clusters escolhido.

Normalização



$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

NORMALIZAÇÃO: MÉTODO DO MÁXIMO E MÍNIMO

Esse método ajusta a escala das variáveis para uma mesma escala (entre 0 e 1), permitindo que nenhuma variável com valores altos afete o modelo mais que variáveis com valores menores.



Alocação de Pesos



ALOCAÇÃO DE PESOS

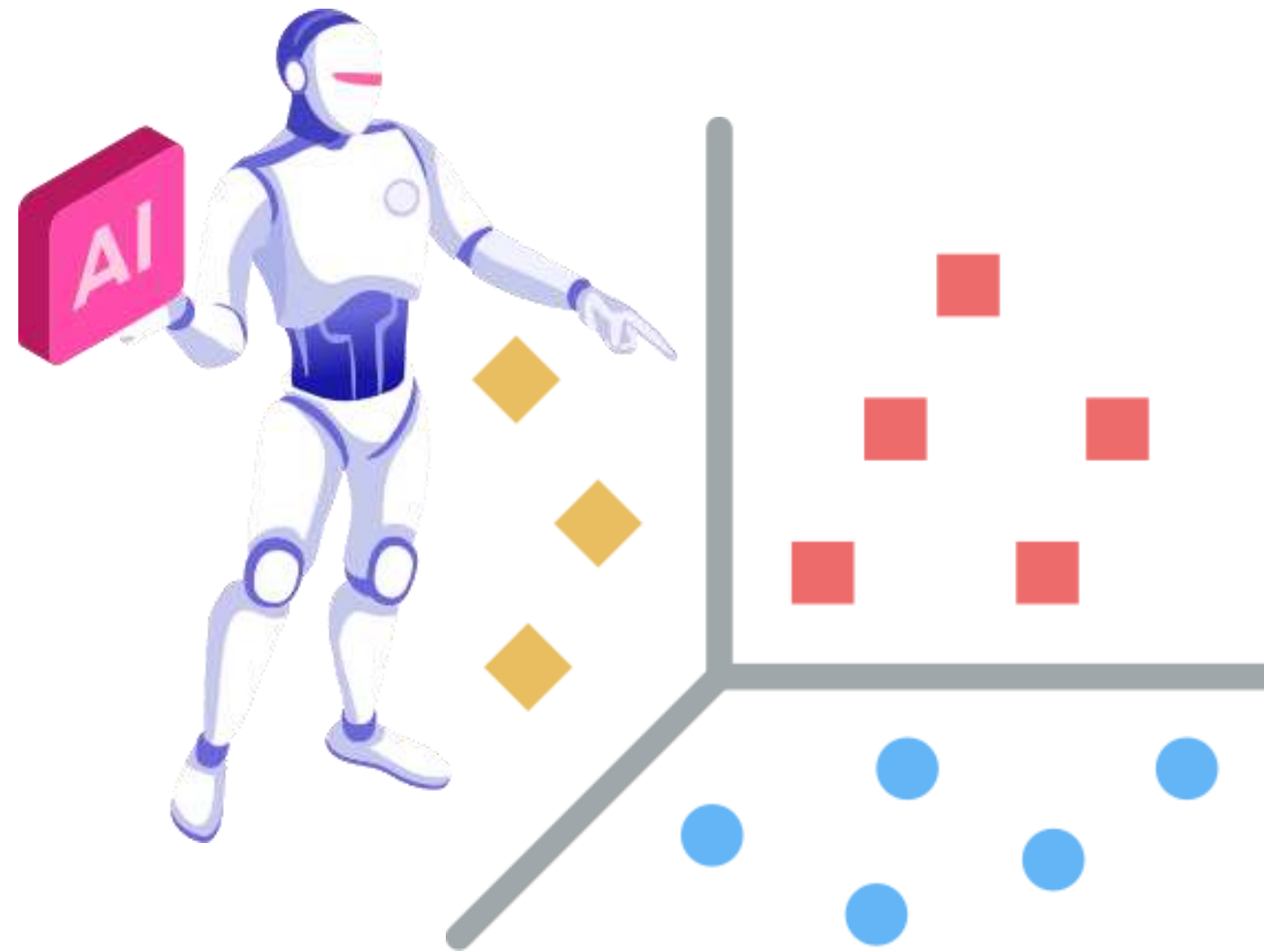
Após a normalização, agora estamos prontos para decidir quais são as principais variáveis que devem influenciar o modelo e escolher valores para os pesos.

Decisão de negócios!!!



Treinamento





TREINAMENTO DO MODELO

No treinamento, o algoritmo analisará todas as variáveis presentes no conjunto de dados a fim de averiguar similaridades nas distâncias dos pontos em relação aos centróides, de acordo com o número de clusters escolhido, e determinará a escolha final de clusters.



Alocação dos Clusters



ALOCAÇÃO DOS CLUSTERS

Treinamento finalizado, temos então um arquivo que pode ser usado para alocar clusters à um conjunto de dados. Pode ser o mesmo que usamos para o treinamento ou dados de clientes novos, caso estivermos lidando com clientes.