

## Algoritmo Pacote de Palavras Virtuais

Nós definimos a classe `PacoteDePalavrasVirtuais` em nosso código. O principal método dela é o de `gerar_dicionario`, apresentado no código a seguir:

```
def gerar_dicionario(self, lista_descritores):  
  
    kmeans = KMeans(n_clusters = QUANTIDADE_PALAVRAS_VIRTUAIS)  
    kmeans = kmeans.fit(lista_descritores)  
    self.dicionario = kmeans.cluster_centers_
```

Neste algoritmo, o dicionário é análogo ao dicionário de palavras do algoritmo original em Processamento de Linguagem Natural. A diferença é que é um dicionário de palavras virtuais gerado pelos centróides do algoritmo `KMeans`.

Por que utilizamos um algoritmo de clusterização para gerar dicionário na classe?

Selecione uma alternativa

- A** Esta é uma das adaptações feitas no algoritmo pela comunidade de visão computacional. Ela é importantíssima nesta fase, pois nós estamos clusterizando os dados que afetarão a performance de classificação e faremos duas classificações na base de dados: uma para poder classificar cada descritor com seu centróide e outra para de fato classificar a imagem como um todo após vetorizar os descritores.
- B** Como são muitos descritores gerados de cada imagem, o dicionário pega os centróides do algoritmo de clusterização com o objetivo de reduzir o tamanho do vocabulário de palavras e assim conseguir melhorar na generalização dos dados pelo histograma gerado e reduzir a complexidade para o algoritmo de aprendizado.
- C** O algoritmo de aprendizado de máquina não suporta dados com tamanha dimensionalidade e não rodaria. Por isso, nós utilizamos os algoritmos de clusterização, para clusterizar os descritores das imagens, pois assim conseguimos reduzir o tamanho do vetor de entrada.