

Full training  
dataset

Stratified  
few-shot  
dataset

1. Random Sampling

2. Semantic Embedding

3. TF-IDF vectors

Prompt

Few-shot  
examples

+

Input  
sentence

LLM  
generates  
completion



Embedding space



TF-IDF space



Free samples



Target / Input sentence



Random sampling

