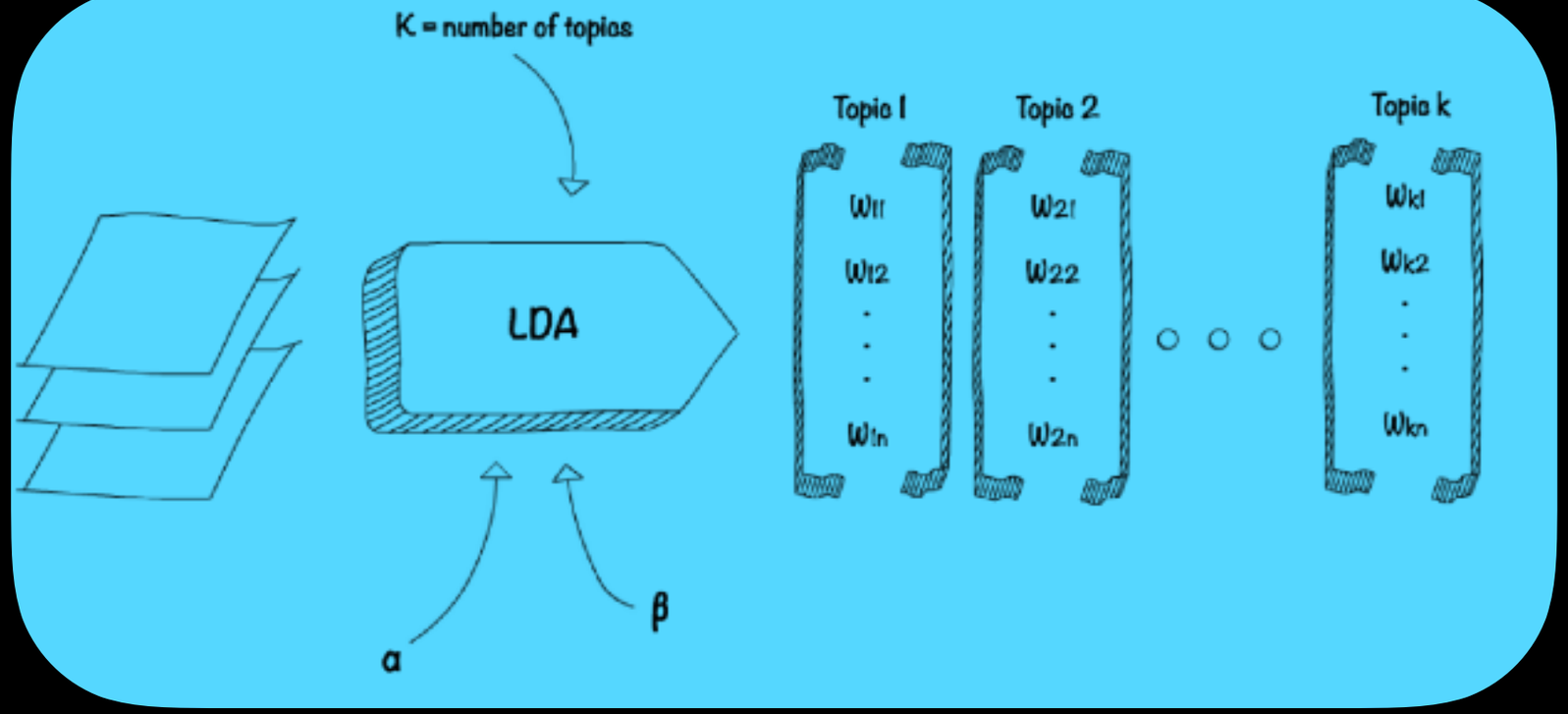

- We **repeat** this process a large number of times (like around 1,000), going through **every term** in **every document** in the corpus, and in each cycle **updating alpha** and **beta**.

- This is the **training process**, which works just like it did for other supervised machine learning techniques.

- However, once we get our trained model, this is just still for some value of **K**, which we don't know is the best value.

- While we don't cover this here, you'll need to repeat this entire process for various values of **K** until we get topics that are distinct from one another.

BY THE END OF THIS SESSION YOU SHOULD BE ABLE TO:

- Explain why lemmatization is a more thorough approach than stemming when normalizing n-grams in a corpus.

- Identify the unigrams, bigrams, and trigrams in a sentence.

- Understand how to calculate a TF-IDF if given the necessary values. (No need to calculate the final product, but know how to set up the calculation.)

- Explain how an LDA model works in terms of alpha and beta.

- Explain how an LDA model works in terms of alpha and beta.
- Understand how to calculate a TF-IDF if given the necessary values. (No need to calculate the final product, but know how to set up the calculation.)