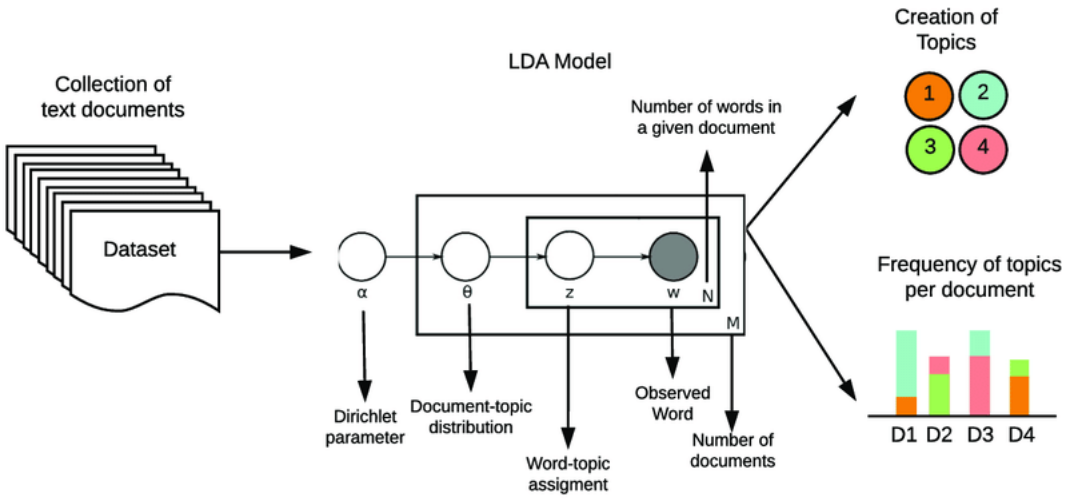
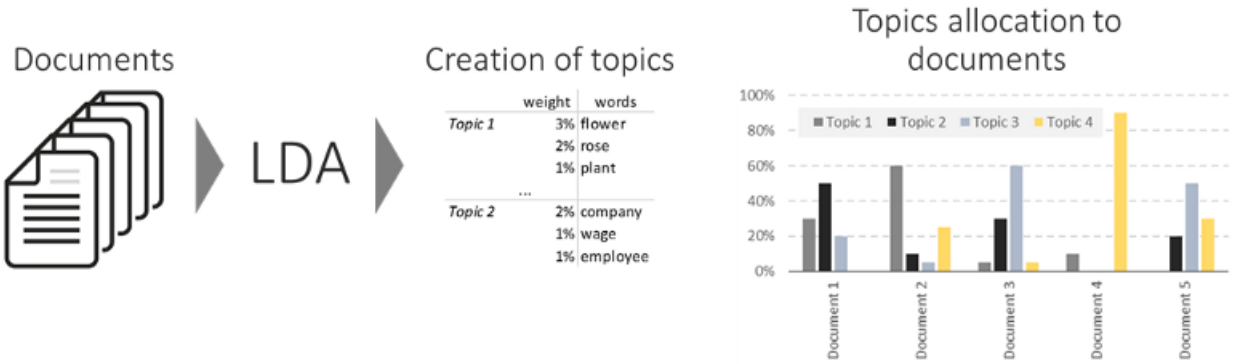
Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: Data Scientist.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

Latent Dirchlet Allocation

In natural language processing, Latent Dirichlet Allocation (LDA) is a Bayesian network (and, therefore, a generative statistical model) that explains a set of observations through unobserved groups, and each group explains why some parts of the data are similar. The LDA is an example of a Bayesian topic model. In this, observations (e.g., words) are collected into documents, and each word's presence is attributable to one of the document's topics. Each document will contain a small number of topics.



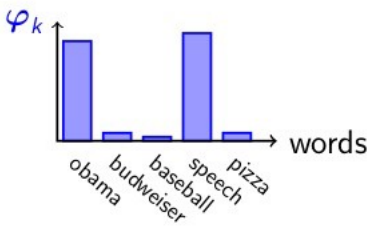


Latent Dirichlet Allocation

LDA discovers topics into a collection of documents.

LDA tags each document with topics.

Topic k



Document d