

Model	Supervised or Unsuper- vised?	Discovery or Measure- ment?	What does it do?
Naive Bayes	Supervised	Measurement	Puts each document into mutually exclusive categories. It tells you the probability that an unseen document is in a particular category given that you have already observed the document.
Fixed base function models	Supervised	Measurement	Predicts what category a particular document fits. Examples of these are Regression models such as OLS, LASSO, Ridge Regression, Perceptron, and Support Vector Machines. Perform best when data are limited and documents contain a lot of independent information about the outcome
Adaptive basis function models	Supervised	Measurement	Similar to fixed base function models. They are good for when you have variables that are informative about the category of interest. Examples of these are classification and regression trees like Random Forest models, Boosting, and Neural Networks.

Clustering	Unsupervised	Discovery	<p>Finds partitions of documents - not existing ones like in supervised models. Each document is one-hot-encoded (or classified strictly as belonging to one category to another, not a mix of categories). Examples of these are K-Means clustering, probabilistic clustering, mixture of multinomials clustering models, mixture of von Mises-Fisher clustering models, mixture of Gaussians models, mixture of Bernoullis clustering models, affinity propagation, hierarchical clustering, and spectral clustering</p>
Topic Models	Unsupervised	Discovery	<p>Similar to clustering models. They are different from clustering models in that they assume documents have proportional membership to all categories rather than being exclusive to one category. The most famous of these models is the Latent Dirichlet Allocation model. There are also structural topic models. With these STM's, you can incorporate "meta data" or information about the document to help identify relevant topics given that information.</p>

Document dings	Embed-	Unsupervised	Discovery	Similar to topic models, but you put the documents into 1-3 categories. These are referred to as low-dimensional space topic models. Examples of these are Principal Component Analysis (PCA), and classic multidimensional scaling.
-------------------	--------	--------------	-----------	--

Source: Damon C. Roberts, 2022