<https://towardsdatascience.com/do-you-know-how-to-choose-the-right-machine-learning-algorithm-among-7-different-types-295d0b0c7f60>

Linear regression: only for continuous variables if no modifications added. Use a loss function such as MAE or MSE. This basic model suffers from its linear structure which prevents it from learning complex patterns. However if the data and task are adapted, this model is powerful.

Logistic regression: for classification: the dependent variable is lo odds of the variable. Basically linear equation passed through sigmoid to get result. Like 1 node of a neural network!

Random Forest: Build a decision tree to get the rules to make the decision. The process contains a lot of randomness and might change at each iteration. Because you only choose between two “rules” at each leaf, intuitively this model often does not capture the entire complexity of a dataset. To solve this issue, we take a subsample of the dataset, a subsample of features, and train a decision tree on it. Repeat the process for many different subsamples. Then you take each tree and decide on the overall decision by “majority” vote.

Neural network: The inputs go through layers which are connected to each other in different fashions (dense, sparse, etc…) and ultimately lead to an output layer which makes the ultimate choice. Based on the result, and a chosen loss function, the model is made to strengthen or weaken certain connections between nodes, which is the basis of the learning process. At each layer, a nonlinear function is required so that the process is not just a linear combination of variables. This use of nonlinearities allows the network to learn complex patterns**.**

Naïve Bayes Classifier for multivariate Bernoulli Models: A naïve Bayes classifier relies on prior information to compute the probability of an event. The Bayes formula relies on the fact that events (features) are independent from one another which is an incredibly strong assumption when you take a look at real world data. Depending on the dataset and the prediction to be made, different variants of Naïve Bayes classifiers can be used.

The Bernouilli naïve bayes classifier takes inputs that are binary, while the multinomial naïve bayes classifier takes inputs that are occurrence counts. In the case of BNB, each feature is assumed to be distributed to a different distribution (multivariate Bernoulli distribution) that is independent from that of other features. To be able to use continuous features you must discretize them first.

The update rule for BNB is as follows:

P(x\_i \mid y) = P(i \mid y) x\_i + (1 - P(i \mid y)) (1 - x\_i)

* C.D. Manning, P. Raghavan and H. Schütze (2008). Introduction to Information Retrieval. Cambridge University Press, pp. 234-265.
* A. McCallum and K. Nigam (1998). [A comparison of event models for Naive Bayes text classification.](http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.1529) Proc. AAAI/ICML-98 Workshop on Learning for Text Categorization, pp. 41-48.
* V. Metsis, I. Androutsopoulos and G. Paliouras (2006). [Spam filtering with Naive Bayes – Which Naive Bayes?](http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.5542) 3rd Conf. on Email and Anti-Spam (CEAS).