Data Mining, CSCI 347, Fall 2019
Naïve Bayes, Sept. 11

Naïve Bayes has worked well in document classification and spam filtering.

Naïve Bayes classifiers

- Require a small amount of training data to estimate the necessary parameters (ratios)
- Are extremely fast compared to more sophisticated methods
- Alleviates the curse of dimensionality
- Decent classifier, bad estimator

Other Naïve Bayes:

- Gaussian Naïve Bayes - likelihood of the features is assumed to be Gaussian (normally distributed)
- Multinomial Naïve Bayes (NMB) – used in text classification where the data are typically represented as word vector counts
- Complement – an adaptation of NMB well suited for imbalanced data sets.
- Bernoulli Naïve Bayes – for classification algorithms for data that is distributed according to multivariate Bernoulli distributions, i.e. there may be multiple features but each one is assumed to be a binary-valued (Bernoulli, Boolean) variable.

From scikit-learn 0.21.3 documentation