Handling Data with Binary features with Naive Bayes assumption:

[redacted]

Naive Bayes assumption assumes independence between all features, for that reason we can use:

$$P(X_1,\ldots,X_N \mid C) = P(X_1 \mid C)\cdots P(X_N \mid C)$$

That is, given the class, the feature probability is calculated independently of the other features and therefore the calculation is not affected by the possible combinations of all features values.

So For example, given a training data with N Binary features, we have $2N$ possible values (True / False for each feature from N features).

Wheres we have 2^N possible combinations of all features values.

Given Naive Bayes assumption, the Naive Bayes classifier don't need to find the relations between features, and it uses all of the training samples to induce on each feature independently, so it will handle well even when **|training samples|** << **|features dimension|.**

In contrast, when Naive Bayes assumption are not given, in order to find the connection between features the ML algorithms (Naive Bayes classifier included) will need an exponential number of training data 2^N, in order to ensure that there are several samples with each combination of features values.

From those reasons, those ML algorithms will need training data such: **|training data|** >> **|features number|**.