Training Test\_with\_stopwords Test\_without\_stopwords

0.9941348973607038 0.997134670487106 0.9914040114613181

**When stop words are removed, accuracy decreases. This is because of the following reasons:**

Stop word is might be the word that has high frequency and which gives meaning to context.

Removing stop words might change meaning of the context, some words may appear "too close" to each other. For e.g spam might look like ham

Generally, for classification ideal way is to remove stop words (in text classification) which do not add meaning to the context. It’s better to keep these words and do some tests with and without them so see how it affects the model. Anyway, you should never remove stop words without thinking about the impact of these words on the problem you are trying to solve.