Remove some n-grams

Let's remove some n-grams from features based on their occurrence frequency in documents of our corpus

- **High frequency n-grams:**
  - Articles, prepositions, etc. (example: and, a, the)
  - They are called **stop-words**, they won't help us to discriminate texts → remove them
- **Low frequency n-grams:**
  - Typos, rare n-grams
  - We don't need them either, otherwise we will likely overfit
- **Medium frequency n-grams:**
  - Those are good n-grams