

- **Masking string fragments for escape sequences:** This involves replacing specific character combinations that have special meanings in text (like \a, \b, \t, \n, which represent bell, backspace, tab, and newline respectively) with some placeholder or removing them. The purpose might be to neutralize these characters so they don't perform their usual functions in a system.
- **Mask all numbers:** Replace all numeric characters in a text with another character or remove them. This is typically done for reasons like privacy or to standardize text for processing.
- **Replace all single-letter words with an empty string:** This involves finding words that consist of a single letter (like 'a' or 'I') and removing them from the text. This might be used to simplify text or focus on more meaningful words in content analysis.
- **Mask email addresses:** Modify email addresses in the text so they are not recognizable or usable, which is often done to protect user privacy or to prevent automated tools from collecting email addresses for spamming purposes.
- Stop words were masked from the dataset
- Lemmatization