

Report

Ques 1: Data Preprocessing

Firstly we read the given dataset line by line and stores in a list then performed preprocessing steps as ordered.

- a. Lowercase the text
- b. Perform tokenization
- c. Remove stopwords
- d. Remove punctuations
- e. Remove blank space tokens

After All we got preprocessed text

2.

Before Preprocessing :

```
1 My 3rd Joyo Pedal, I'm falling in love with that company, solid great
sounding pedals for a fraction of other brands.
2 The cover is very beautiful, mainly the interior. One con I find is that
it is too heavy and uncomfortable to transport.
3 This guitar is perfect in every way! It's so easy to play and sounds
fantastic! Perfectly balanced with zero neck dive. The bomb inlays just
look cool! I definitely plan on buying more Hardluck Kings guitars!
4 Beautiful, well-made strap. Matches my guitar nicely. Perfect.
5 Amazing and highly different Wah from ibanez,This is the WAH YA
WANT!!,too many reasons!,Just check it out!!!
```

After Preprocessing :

```
1 ['3rd', 'joyo', 'pedal', "'m", 'falling', 'love', 'company', 'solid',
'great', 'sounding', 'pedals', 'fraction', 'brands']
2 ['cover', 'beautiful', 'mainly', 'interior', 'one', 'con', 'find',
'heavy', 'uncomfortable', 'transport']
3 ['guitar', 'perfect', 'every', 'way', "'s", 'easy', 'play', 'sounds',
'fantastic', 'perfectly', 'balanced', 'zero', 'neck', 'dive', 'bomb',
'inlays', 'look', 'cool', 'definitely', 'plan', 'buying', 'hardluck',
'kings', 'guitars']
4 ['beautiful', 'well-made', 'strap', 'matches', 'guitar', 'nicely',
'perfect']
5 ['amazing', 'highly', 'different', 'wah', 'ibanez', 'wah', 'ya', 'want',
'many', 'reasons', 'check']
```

Ques 2:

Firstly, we read the data from the given dataset line by line and stored in the list named as docs then build the inverted index of the docs, save as .pkl file and then load the inverted index to further processing of the input queries and finally Output will be generated respectively.