

Assignment 4 Analysis

Q2: Prediction.

Analysis :

In Q2 , the prediction varies with the flags lemmatized and no_stopwords as follow:

- If none is set , the prediction percentage is 63
- If lemmatized is set , the prediction percentage is 78
- If no_stopword is set , the prediction percentage is 63
- If both are set , the prediction percentage is 79

As seen above , lemmatization and removing stop words from the text gives higher prediction percentage , i.e. number of correct predictions as compared to the original data. Though only removing stop words without lemmatization has no effect on the prediction percentage , which means lemmatization has a very strong effect on the prediction model.

Q3 : Precision

Precision for the 4 cases was:

1. 0.552 , both not set
2. 0.519 , lemmatized set
3. 0.539 , no_stopword set
4. 0.519 , both set

Lemmatization has inverse effect on precision .