From sklearn.feature\_extraction.text import CountVectorizer

>>> corpus = [

... 'This is the first document.',

... 'This document is the second document.',

... 'And this is the third one.',

... 'Is this the first document?',

... ]

>>> vectorizer = CountVectorizer()

>>> print(vectorizer.get\_feature\_names())

['and', 'document', 'first', 'is', 'one', 'second', 'the', 'third', 'this']

>>> X = vectorizer.fit\_transform(corpus)

>>> print(X.toarray())

[[0 1 1 1 0 0 1 0 1]

[0 2 0 1 0 1 1 0 1]

[1 0 0 1 1 0 1 1 1]

[0 1 1 1 0 0 1 0 1]]

**fit(raw\_documents, y=None)[source]**

Learn a vocabulary dictionary of all tokens in the raw documents.

**fit\_transform(raw\_documents, y=None)[source]**

Learn the vocabulary dictionary and return term-document matrix.

**get\_feature\_names()[source]**

Array mapping from feature integer indices to feature name

**Accuracy-** .Accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition:

Accuracy=Number of correct predictions/Total number of predictions

**precision** (also called [positive predictive value](https://en.wikipedia.org/wiki/Positive_predictive_value)) is the fraction of relevant instances among the retrieved instances

**Recall** (also known as [sensitivity](https://en.wikipedia.org/wiki/Sensitivity_and_specificity)) is the fraction of relevant instances that have been retrieved over the total amount of relevant instances.

Total documents- 40

Total Positive doc- 30

Total Negative doc -10

Total Predicted true- 25

Total Predicted correctly 20

Precison is 20/25

Recall is 20/40

When a [search engine](https://en.wikipedia.org/wiki/Search_engine_(computing)) returns 30 pages only 20 of which were relevant while failing to return 40 additional relevant pages.

Precision is 20/30 = 2/3

Recall is 20/60 = 1/3.

So, in this case, precision is "how useful the search results are", and recall is "how complete the results are".



Accuracy=TP+TN / (TP+FP+TN+FN)

Precision = TP/(TP + FP)

Recall = TP/(TP+FN)