

Tf-Idf example

In [1]:

```
1 from sklearn.feature_extraction.text import TfidfVectorizer
2 import pandas as pd
3 texts = [
4     "good movie", "not a good movie", "did not like",
5     "i like it", "good one"
6 ]
7 # using default tokenizer in TfidfVectorizer
8 tfidf = TfidfVectorizer(min_df=2, max_df=0.5, ngram_range=(1, 2))
9 features = tfidf.fit_transform(texts)
10 pd.DataFrame(
11     features.todense(),
12     columns=tfidf.get_feature_names()
13 )
```

Out[1]:

	good movie	like	movie	not
0	0.707107	0.000000	0.707107	0.000000
1	0.577350	0.000000	0.577350	0.577350
2	0.000000	0.707107	0.000000	0.707107
3	0.000000	1.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000

In []:

1