

Access vocabulary learned by CountVectorizer

Vocab = CV. vocabulary.get_keys()

(cosine-similarity)

cos-sim = cosine-similarity (count-matrix)

vector of
combined features

Day 42 William Kornit

Jan 10

What we are doing!

Data is made of alphabets "string". ML
works better with number.

Goal 1: Identify all features of interest
and put all in one column in
the df. and fill the null ~~rows~~ rows

Goal 2: Convert that column of features
into integer. using countvectorizer and fit_transform

Goal 3: Calculate the similarity of ~~each~~
df

Calculate 4: Cosine similarity given the
similarity of movies based on their index.
So to access we need to define a function
do that.

```
def get_index_from_title(title):  
    return df[df.title == title]["index"].values[0]
```

Wh