

**Evaluating a dish**

Some codes are meant to transform features- normalise numericals or turn text into vectors, or fill up missing data, they are **transformers**. Other codes are meant to predict variables by fitting an algorithm such as a random forest or support vector machine (SVM), they are **estimators.**

So, in a pipeline, we first sequentially apply a list of transformers (data modelling) and then a final estimator (ML model).

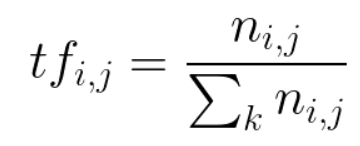
The transform steps must implement fit() and transform().

The final step, estimator, should implement fit() and predict().

**Transformer used if TFIDF**

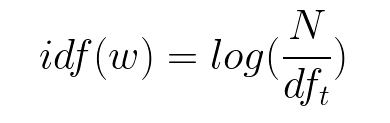
**Term Frequency (TF)**

The number of times a word appears in a document divided by the total number of words in the document. Every document has its own term frequency.

****

**Inverse Data Frequency (IDF)**

The log of the number of documents divided by the number of documents that contain the word ***w***. Inverse data frequency determines the weight of rare words across all documents in the corpus.



Multinomial NB Classifier : Estimator

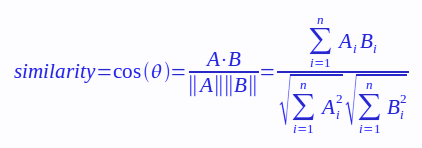
Preferences for few dishes are known.

For each such dish, the ingredient list gives a particular rating.

We need a rating of x dish with y list of ingredients.

**Recommendation**

Cosine Similarity between highly rated dishes and not rated dishes is used to find best recommendations

****