[lr:tokenization]Tokenization
[lr:wordrep]Words Representaion
[lr:ml]Machine Learning
[lr:oversampling]Oversampling
[lr:lm]Language Models
[lr:evalmetrics]Evaluation Metrics
1
2
Hazm
NLTK
3
?
?
?
?
?
4
?
Trem frequency (TF):
TF
TF
??

$$tf\,(t,D) = \frac{f_{t,d}}{\Sigma_{t^{`}\in d} f_{t^{`},d}} \quad (1)$$

Inverse Document Frequency (iDF):
??

$$idf\,(t,D) = \log\frac{N}{|\{d\in D, t\in d\}|} \quad (2)$$

TF
iDF
??

$$tf - idf\,(t,d,D) = tf\,(t,D)\,.idf\,(t,D) \quad (3)$$

?
5
?
?
Gaussian Naive Bayes
SVC
LinearSVC
Random Forest
Logistic Regression
?
6

$$P\,(y|X) = \frac{P(X|y).P(X)}{P(y)}$$

X
y
$P\,(X|y)$
y

$$p\,(x=v|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}e^{-\frac{(\epsilon-\mu_k)^2}{2\mu_k^2}}$$