

data pre processing

why is it important?



lightning talk

VOSE CAIU NAS TRAQUINAGEM DO
RELAMPAGO MARQUINHOS



lightning talk

VOSE CAIU NAS TRAQUINAGEM DO
RELAMPAGO MARQUINHOS





mikaeri ohana

why
should I
do that?

depends on
the
domain

case
align
ment





tokenization



stop
words



stem
ming



lemma
tization



bag
of
words

charmander is the best

1

is

0

squirtle

0

are

1

charmander

1

best

0

fire

1

the

In [23]: `from sklearn.feature_extraction.text import CountVectorizer`



libra
ries



material

<https://scikit-learn.org/stable/modules/preprocessing.html>

<https://hackernoon.com/what-steps-should-one-take-while-doing-data-preprocessing-502c993e1caa>

<https://towardsdatascience.com/data-pre-processing-techniques-you-should-know-8954662716d6>

<https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/>

<https://www.ques10.com/p/17025/discuss-different-steps-involved-in-data-preproces/>

<https://www.techopedia.com/definition/14650/data-preprocessing>

<https://www.datacamp.com/community/tutorials/stemming-lemmatization-python>



thank you!



@miohana



@mikaeriohana



Mikaeri Ohana

