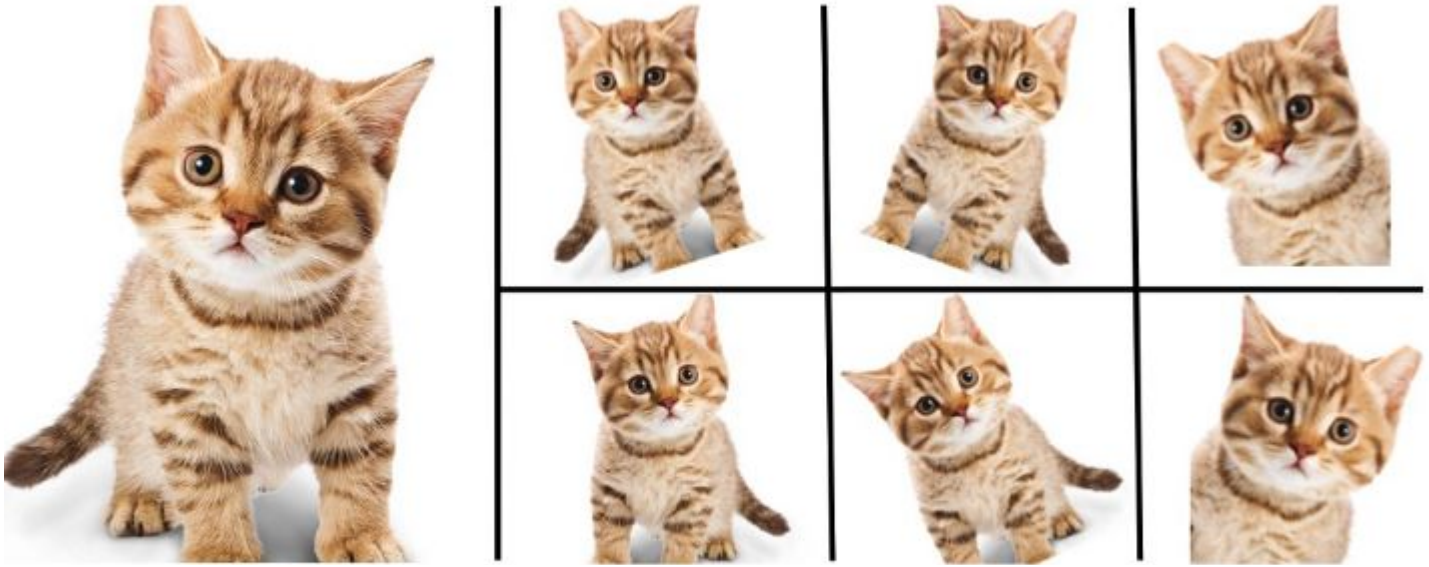
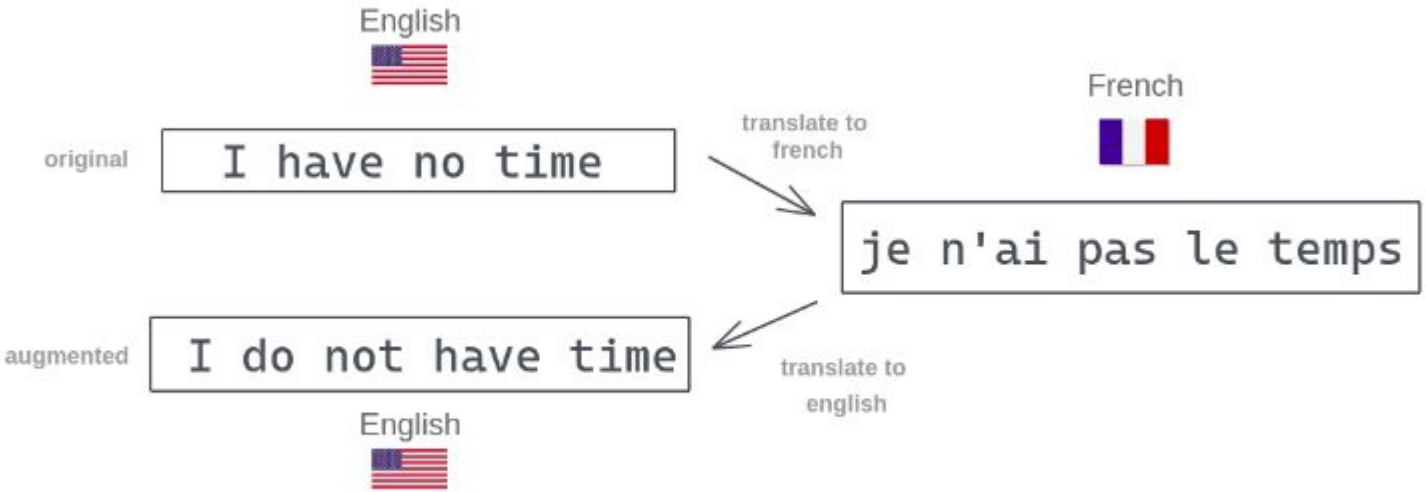
# Text augmentation

# O que é augmentation?



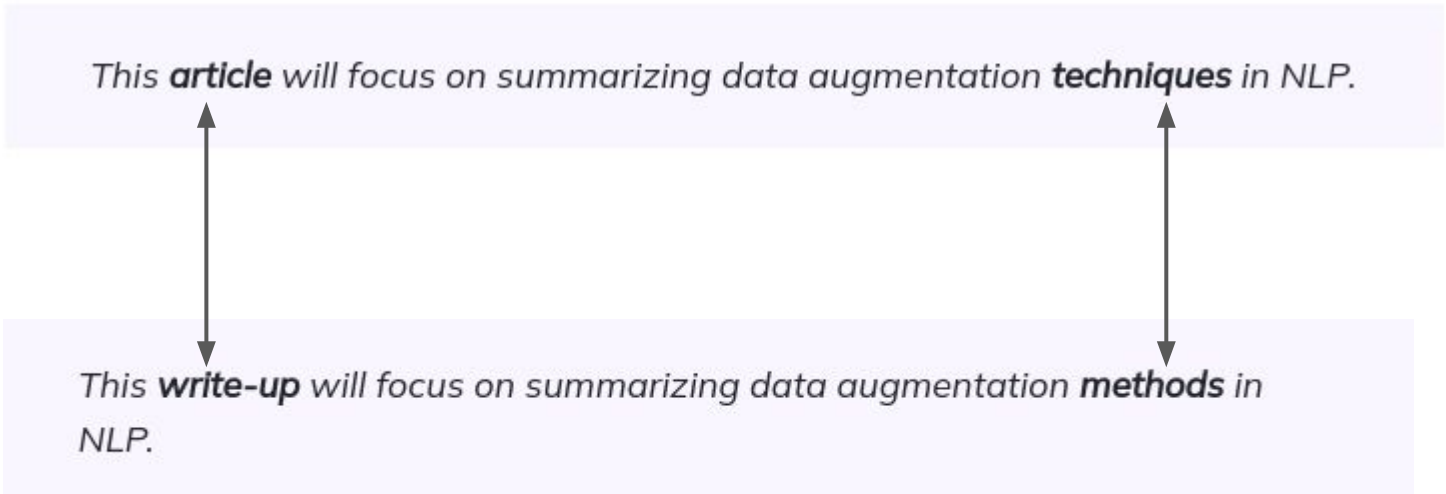
# Techniques

# Back translation




*Amit Chaudhary "Back Translation for Text Augmentation with Google Sheets"*

# Synonym Replacement



# Random Insertion

*This **article** will focus on summarizing data augmentation **techniques** in NLP.*

*This article will focus on **write-up** summarizing data augmentation techniques in NLP **methods**.*

# Random Swap

*This **article** will focus on summarizing data augmentation **techniques** in NLP.*

*This **techniques** will focus on summarizing data augmentation **article** in NLP.*

# Random Deletion

*This **article** will focus on summarizing data augmentation **techniques** in NLP.*

*This article will focus on **write-up** summarizing data augmentation techniques in NLP **methods**.*

# Tools

# NLPAug Library

NLPAug offers three types of augmentation:

- Character level augmentation
- Word level augmentation
- Sentence level augmentation

In each of these levels, NLPAug provides all the methods discussed in the previous sections such as:

- random deletion,
- random insertion,
- shuffling,
- synonym replacement,
- etc.

# Bests Practices

- Do not validate using the augmented data.
- If you're doing K-fold cross-validation, always keep the original sample and augmented sample in the same fold to avoid overfitting.
- Always try different augmentation approaches and check which works better.
- A mix of different augmentation methods is also appreciated but don't overdo it.
- Experiment to determine the optimal number of samples to be augmented to get the best results.
- Keep in mind that data augmentation in NLP does not always help to improve model performance.

# References


- Augment to Prevent: Short-Text Data Augmentation in Deep Learning for Hate-Speech Classification - https://dl.acm.org/doi/10.1145/3357384.3358040
- EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks - https://arxiv.org/abs/1901.11196
- A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios - https://arxiv.org/abs/2010.12309