



PROCESSAMENTO DE LINGUAGEM NATURAL ENRIQUECIMENTO DE DADOS

Enriquecimento de Dados

Envolve a criação de
variações de dados
textuais para enriquecer
conjuntos de treinamento





Motivação

- 01 Conjuntos de dados limitados
- 02 Melhoria da generalização do modelo
- 03 Balanceamento de dados
- 04 Robustez a ruídos e erros
- 05 Adaptação a diferentes dialetos
- 06 Teste e avaliação de modelos



Motivação

- 01 Conjuntos de dados limitados
- 02 Melhoria da generalização do modelo
- 03 Balanceamento de dados
- 04 Robustez a ruídos e erros
- 05 Adaptação a diferentes dialetos
- 06 Teste e avaliação de modelos



Motivação

- 01 Conjuntos de dados limitados
- 02 Melhoria da generalização do modelo
- 03 Balanceamento de dados
- 04 Robustez a ruídos e erros
- 05 Adaptação a diferentes dialetos
- 06 Teste e avaliação de modelos



Motivação

- 01 Conjuntos de dados limitados
- 02 Melhoria da generalização do modelo
- 03 Balanceamento de dados
- 04 Robustez a ruídos e erros
- 05 Adaptação a diferentes dialetos
- 06 Teste e avaliação de modelos



Motivação

- 01 Conjuntos de dados limitados
 - 02 Melhoria da generalização do modelo
 - 03 Balanceamento de dados
 - 04 Robustez a ruídos e erros
 - 05 Adaptação a diferentes dialetos
 - 06 Teste e avaliação de modelos
-



Motivação

- 01 Conjuntos de dados limitados
- 02 Melhoria da generalização do modelo
- 03 Balanceamento de dados
- 04 Robustez a ruídos e erros
- 05 Adaptação a diferentes dialetos
- 06 Teste e avaliação de modelos

Enriquecimento de Dados



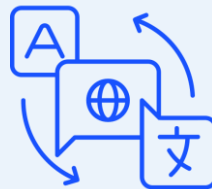
Exemplos de técnicas



Substituição por
sinônimos



Inserção de
palavras



Tradução reversa



Remoção de
palavras

Substituição por Sinônimos



Ela comprou um carro novo



Ela adquiriu um veículo novo

Inserção de palavras



O livro está na mesa.



O livro **interessante** está na mesa.

Tradução reversa



A tecnologia está avançando rapidamente.

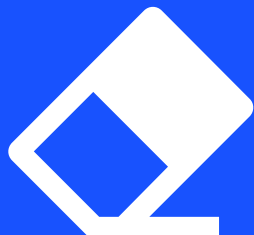


Technology is advancing rapidly. 🇬🇧



A tecnologia progride rapidamente.

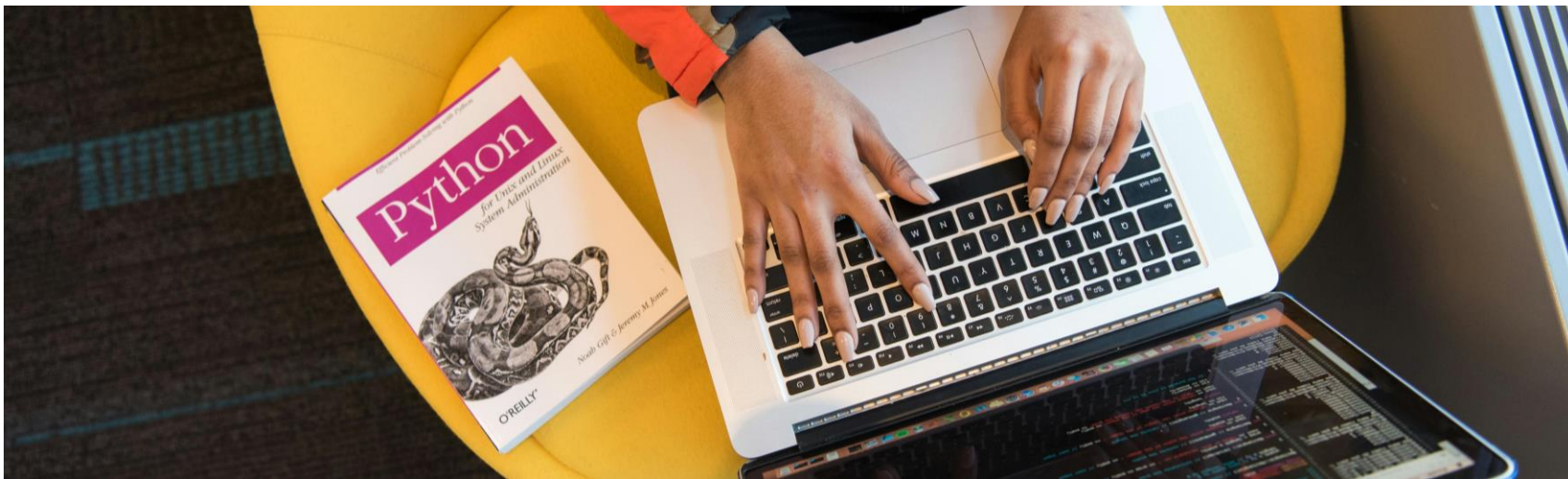
Remoção de palavras



Ele estava jogando **bola** no parque.



Ele estava jogando no parque.



NLPaug



nlpaug

```
$ pip install numpy requests nlpaug
```

```
01 import nlpaug.augmenter.word as naw
```

```
02 a = naw.SynonymAug(  
    aug_src='wordnet', lang='por')
```

```
03 s = a.augment(texto)
```



nlpaug

```
$ pip install numpy requests nlpaug
```

```
01 import nlpaug.augmenter.word as naw
```

```
02 a = naw.SynonymAug(  
    aug_src='wordnet', lang='por')
```

```
03 s = a.augment(texto)
```



nlpaug

```
$ pip install numpy requests nlpaug  
  
01 import nlpaug.augmenter.word as naw  
  
02 a = naw.SynonymAug(  
    aug_src='wordnet', lang='por')  
  
03 s = a.augment(texto)
```



nlpaug

```
$ pip install numpy requests nlpaug  
  
01 import nlpaug.augmenter.word as naw  
  
02 a = naw.SynonymAug(  
    aug_src='wordnet', lang='por')  
  
03 s = a.augment(texto)
```



PUC Minas
Virtual