

Processamento de Língua Natural (PLN)

Willian Massami Watanabe

Processamento de Língua Natural (PLN)

Natural language processing (NLP) uses machine learning to reveal the structure and meaning of text. With natural language processing applications, organizations can analyze text and extract information about people, places, and events to better understand social media sentiment and customer conversations.

<https://cloud.google.com/learn/what-is-natural-language-processing>

Processamento de Língua Natural (PLN)

Natural language processing (NLP) uses machine learning to reveal the structure and meaning of text. With natural language processing applications, organizations can analyze text and extract information about people, places, and events to better understand social media sentiment and customer conversations.

<https://cloud.google.com/learn/what-is-natural-language-processing>

Natural language processing (NLP) is a machine learning technology that gives computers the ability to interpret, manipulate, and comprehend human language.

<https://aws.amazon.com/what-is/nlp/>

Processamento de Língua Natural (PLN)

Natural language processing (NLP) uses machine learning to reveal the structure and meaning of text. With natural language processing applications, organizations can analyze text and extract information about people, places, and events to better understand social media sentiment and customer conversations.

<https://cloud.google.com/learn/what-is-natural-language-processing>

Natural language processing (NLP) is a machine learning technology that gives computers the ability to interpret, manipulate, and comprehend human language.

<https://aws.amazon.com/what-is/nlp/>

Natural language processing, or NLP, combines computational linguistics—rule-based modeling of human language—with statistical and machine learning models to enable computers and digital devices to recognize, understand and generate text and speech.

<https://www.ibm.com/topics/natural-language-processing>



...



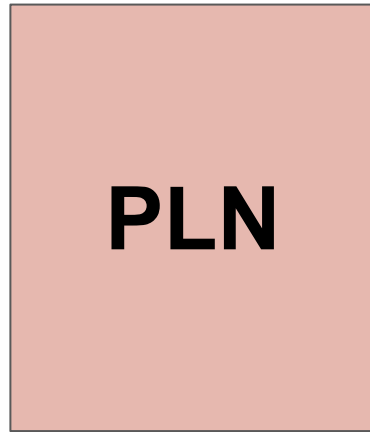
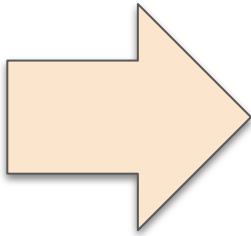
Textos



...



Textos

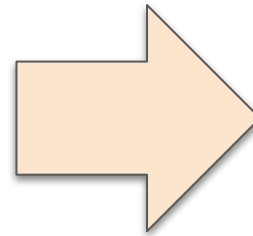
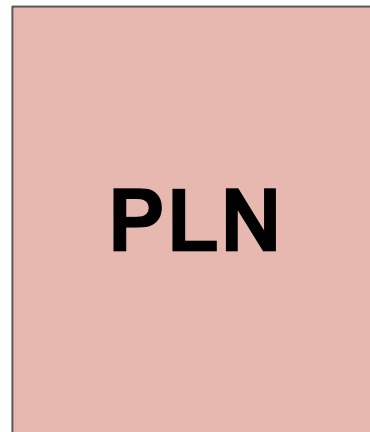
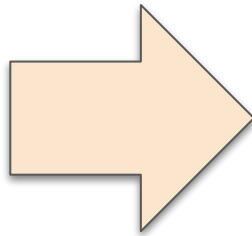




...

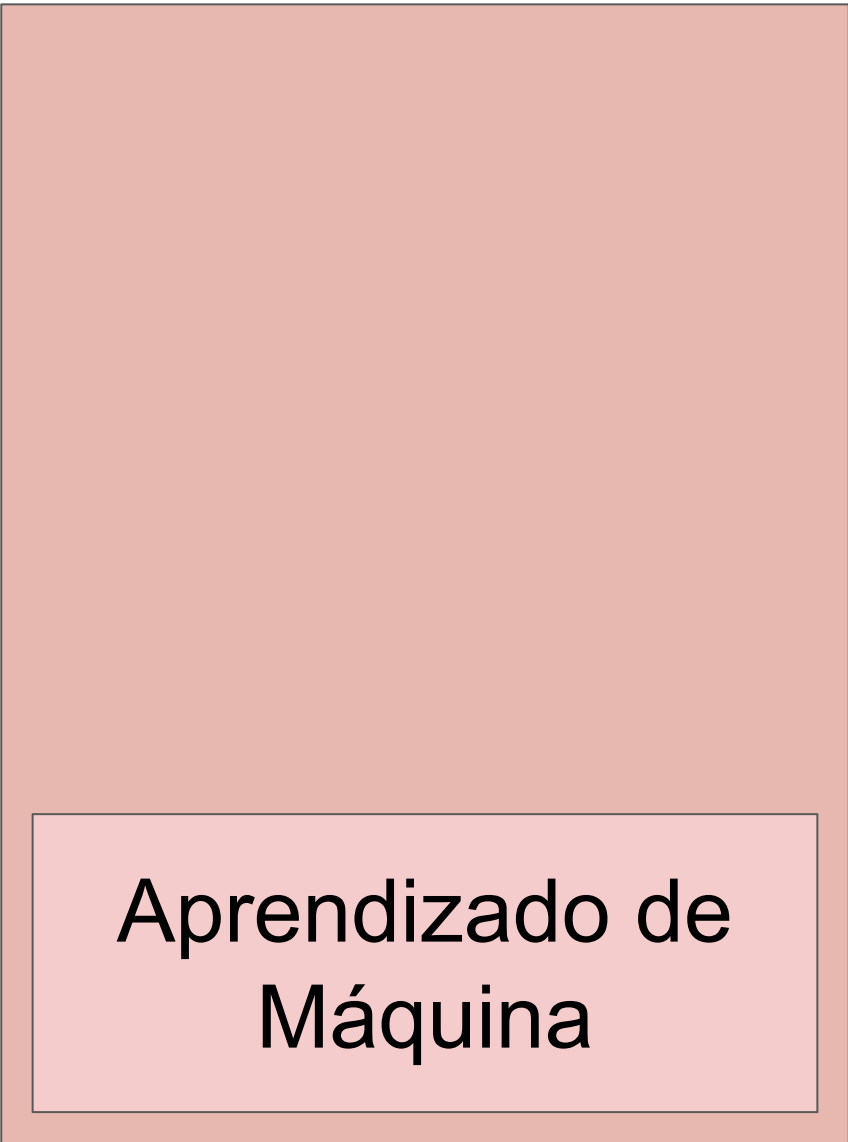


Textos



Aplicações

- Análise de sentimento
- Classificação
- Geração de texto
- Análise de documentos
- Recuperação de informação
- Sumarização
- entre outras



Aprendizado de Máquina

PLN

Aprendizado de máquina

- Aprendizado supervisionado
- Aprendizado não supervisionado
- Aprendizado profundo

A large light red rectangle contains two smaller rectangles. The top rectangle is light blue and contains the text 'Pré-processamento'. The bottom rectangle is light pink and contains the text 'Aprendizado de Máquina'.

Pré-processamento

**Aprendizado de
Máquina**

PLN

Pré-processamento

- Tokenization
- Part-Of-Speech
- Named Entities Recognition
- ...

Aprendizado de máquina

- Aprendizado supervisionado
- Aprendizado não supervisionado
- Aprendizado profundo

A diagram showing the components of PLN (Processamento de Linguagem Natural) stacked vertically within a large light red container. The top two components, 'Pré-processamento' and 'Extração de características', are in light blue boxes. The bottom component, 'Aprendizado de Máquina', is in a light red box. The entire stack is labeled 'PLN' at the bottom.

Pré-processamento

**Extração de
características**

**Aprendizado de
Máquina**

PLN

Pré-processamento

- Tokenization
- Part-Of-Speech
- Named Entities Recognition
- ...

Bag-Of-Words

N-Gram

Aprendizado de máquina

- Aprendizado supervisionado
- Aprendizado não supervisionado
- Aprendizado profundo

Pré-processamento

Tarefas de PLN

- Tokenization
- Part-Of-Speech
- Named Entities Recognition
- ...

Roger Federer é um jogador muito bom.

Pré-processamento

Tarefas de PLN

- Tokenization
- Part-Of-Speech
- Named Entities Recognition
- ...

Roger|Federer|é|um|jogador|muito|bom.

Pré-processamento

Tarefas de PLN

- Tokenization
- Part-Of-Speech
- Named Entities Recognition
- ...

SUJEITO

VERBO

PREDICATIVO

Roger|Federer|é|um|jogador|muito|bom.

Pré-processamento

Tarefas de PLN

- Tokenization
- Part-Of-Speech
- Named Entities Recognition
- ...

SUJEITO

VERBO

PREDICATIVO

Roger|Federer|é|um|jogador|muito|bom.



ENTIDADE

Roger Federer (Basileia, 8 de agosto de 1981) é um ex-tenista suíço que já foi recordista de títulos de Grand Slam com vinte conquistas.

Pré-processamento

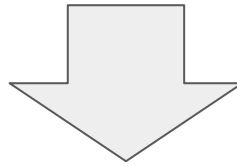
**Extração de
características**

Texto Roger Federer é um jogador muito bom.

Pré-processamento

Extração de
características

Texto Roger Federer é um jogador muito bom.



**Vetor de
Características**

[0 1 1 0 0 1 1 0 0 1 1 1]

Pré-processamento

Extração de
características

Texto Roger Federer é um jogador muito bom.

**Vetor de
Características**

[0 1 1 0 0 1 1 0 0 1 1 1]

Aprendizado de Máquina

Pré-processamento

Extração de
características

Bag-Of-Words

Sequências

Texto Roger Federer é um jogador muito bom.

**Vetor de
Características**

[0 1 1 0 0 1 1 0 0 1 1 1]

Aprendizado de Máquina

Aplicações

- Classificação
- Geração de texto
- Recuperação de informação
- entre outras

Pré-processamento

Extração de
características

Texto Roger Federer é um jogador muito bom.

**Vetor de
Características**

[0 1 1 0 0 1 1 0 0 1 1 1]

Aprendizado de Máquina

Pré-processamento

Pré-processamento



- Tokenization
- Part-Of-Speech
- Named Entities Recognition
- ...

Pré-processamento

Pré-processamento



- Tokenization
- Part-Of-Speech
- ...

Extração de
características

Aprendizado de
Máquina

Bag-Of-Words



Aplicações

- Recuperação de Informação
- Classificação de documentos
- Topicalização

Pré-processamento

Pré-processamento



- Tokenization
- Part-Of-Speech
- ...

Extração de
características

Aprendizado de
Máquina

Bag-Of-Words



Aplicações

- Recuperação de Informação
- Classificação de documentos
- Topicalização

Extração de
características

Aprendizado
profundo

Deep learning



Aplicações

- Classificação de documentos
- Geração de textos

Semana 1 - Tarefas de PLN

Pré-processamento



- Tokenization
- Part-Of-Speech
- ...

Semana 2 - Modelo Bag-Of-Words

Semana 3 - Classificação de documentos com BOW

Semana 4 - Redução de dimensionalidade e ngrams

Semana 5 - Topicalização

Bag-Of-Words



Aplicações

- Recuperação de Informação
- Classificação de documentos
- Topicalização

Semana 6 - Word Embeddings

Semana 7 - Deep Learning

Semana 8 - Geração de Textos

Semana 9 - Transformers

Deep learning



Aplicações

- Classificação de documentos
- Geração de textos