

MBAUSP
ESALQ

TEXT MINING, SENTIMENT ANALYSIS E NLP

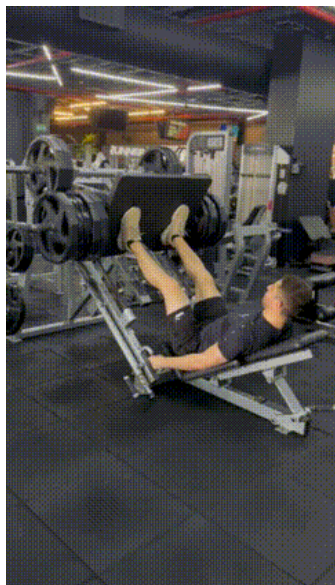
Professor Guilherme Lima

MBAUSP ESALQ

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.

Proibida a reprodução, total ou parcial, sem autorização.

Lei nº 9610/98



X

Minhas redes

guilhermelimadev



guilherme-lima-developer



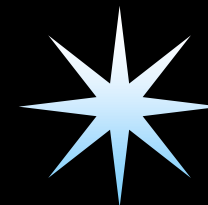
01

Comunicação e uma linguagem

Fundamental para o text mining

Uma língua não é
apenas um conjunto
de palavras.
Ela representa a
cultura, as tradições e
a união de um povo.

Carrega a **história** e a **identidade** de uma comunidade, tornando-se parte essencial do que ela é.



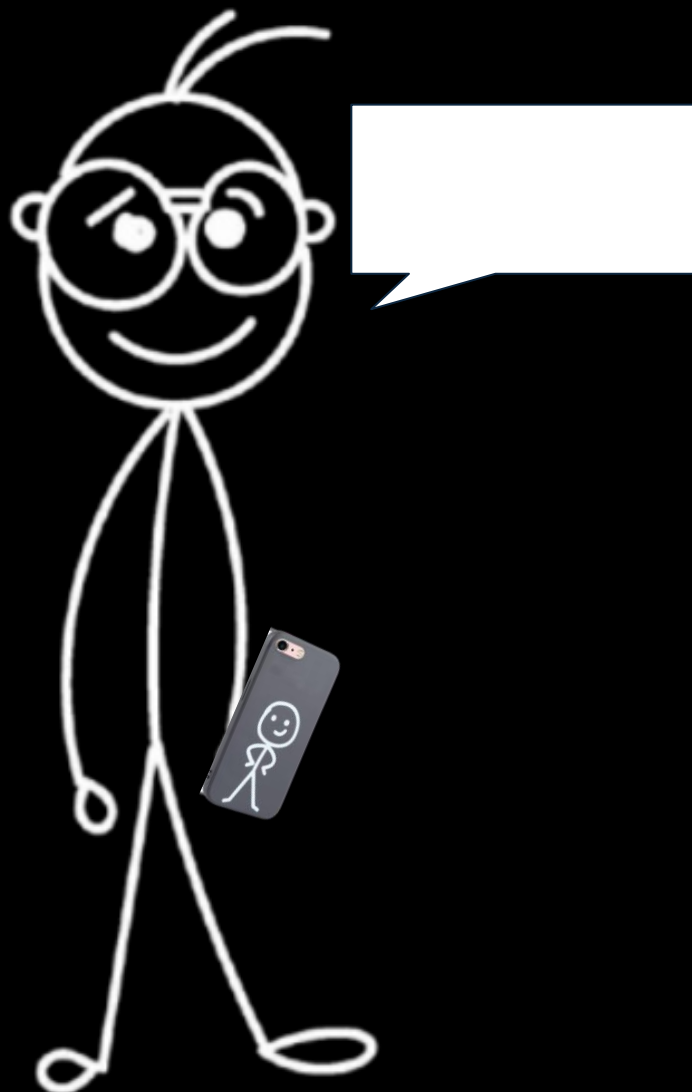
Contextualizar



Valério

Um cientista de dados em uma empresa







~~Linguagem de
programação~~



Linguagem natural



Binário



Algumas áreas da ciência da computação...

NPL

Processamento de
Linguagem Natural

Análise de
Sentimento

Identificação de
emoções textuais

LLMs

Redes neurais para
texto

Text Mining

Extração de
informações textuais

Algumas áreas da ciência da computação...



NLP – Processa e entende a linguagem humana por meio de algoritmos computacionais.

Análise de Sentimento – Detecta emoções e opiniões em textos, classificando como positivas, negativas ou neutras.

LLMs – Modelos de IA treinados em grandes volumes de texto para gerar e compreender linguagem natural.

Text Mining – Extrai padrões, insights e informações úteis a partir de grandes quantidades de texto.



Text Mining

Extrair informações valiosas de textos



Sobrecarga de informações

Vivemos em uma era de sobrecarga de informações.

A maior parte dos dados gerados está em formato **textual**, seja em redes sociais, documentos empresariais ou artigos científicos.





A maior parte dos dados gerados está em formato textual, seja em redes sociais, documentos empresariais ou artigos científicos.



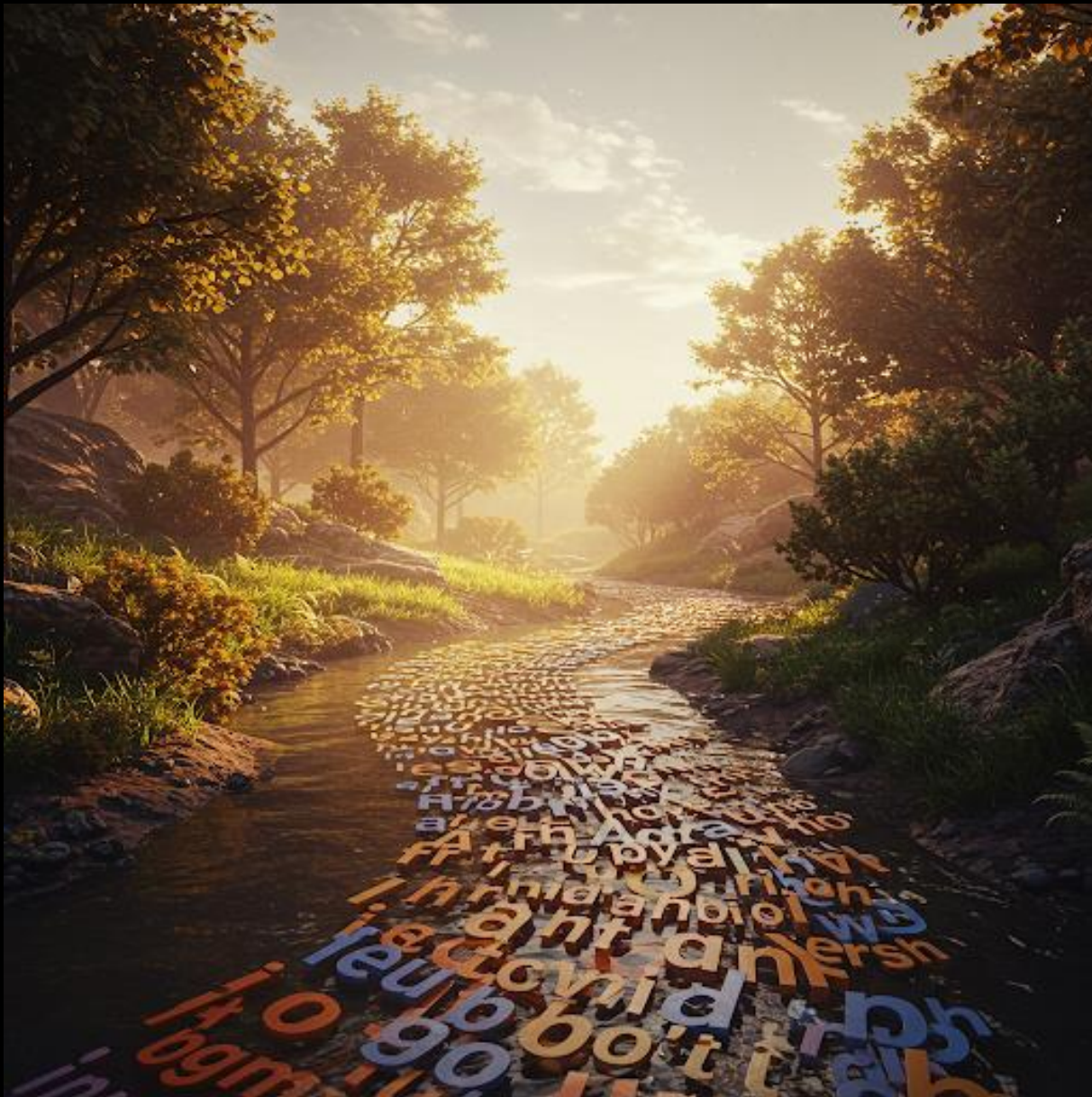
Mas como extrair conhecimento desse volume gigantesco de textos?
Text Mining surge como a solução para extrair *valor* e *conhecimento* desses dados.

→ Desvendando o Text Mining

Processo de descobrir padrões, informações úteis e conhecimento a partir de grandes volumes de dados textuais.



Paulo Silva Gomes 924935.991-87



Garimpar
ouro em
meio a um
rio de texto.

→ Onde o Text Mining Brilha?

1

Análise de Sentimentos

2

Extração de Informação

3

Sumarização de Textos

4

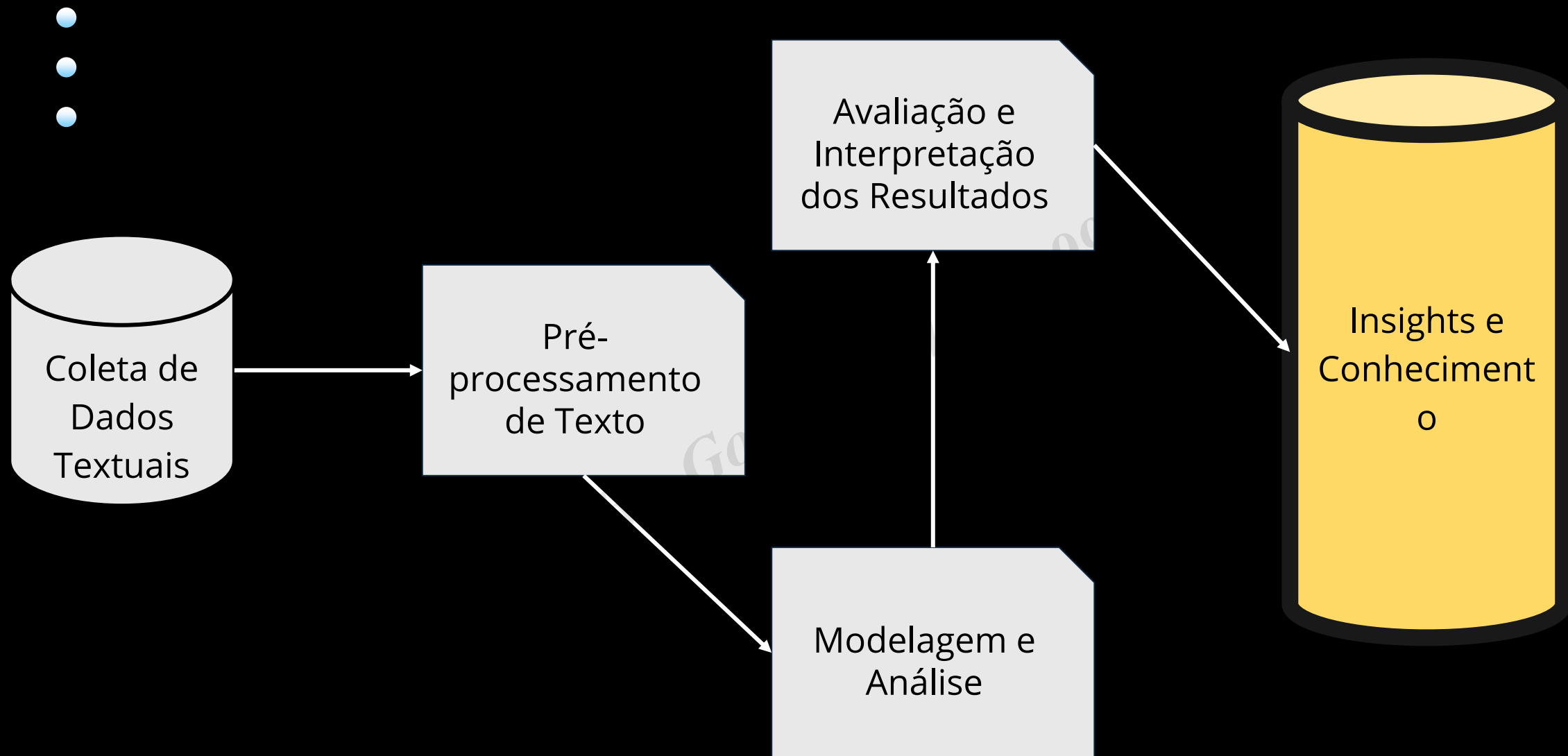
Classificação de Textos

5

Modelagem de Tópicos

6

Detecção de Fraudes e Spam



Pré-processamento de Texto

Paulo Silva Gomes 924.935.991-87

Pré-processamento de Texto

Lowercasing

Padronizar o texto.

Texto original: "Eu amo DS!"

Após lowercasing: ["eu", "amo", "ds", "!"]

Paulo Silva Gomes 924.935.991-87

Pré-processamento de Texto

Lowercasing

Padronizar o texto.

Texto original: "Eu amo DS!"

Após lowercasing: ["eu", "amo", "ds", "!"]

Remoção de Pontuação

Eliminar sinais de pontuação.

Texto original: "Eu amo DS!"

Após remoção: ["eu", "amo", "ds"]

Pré-processamento de Texto

Tokenização

Dividir um texto em unidades menores

Texto original: "Eu amo MBA!"

Após tokenização: ["Eu", "amo", "MBA", "!"]

Paulo Silva Gomes 924.935.991-87

Pré-processamento de Texto

Tokenização

Dividir um texto em unidades menores

Texto original: "Eu amo DS!"

Após tokenização: ["Eu", "amo", "DS", "!"]

Stopwords

palavras comuns que não agregam muito significado ao conteúdo

Exemplo: "a", "o", "e", "de", "que", "em"

Pré-processamento de Texto

Stemming

Reduzir palavras à sua raiz

Texto original: "Eu amo DS!"

Após lowercasing: Eu am DS

Paulo Silva Gomes 924.935.991-87

Pré-processamento de Texto

Stemming

Reduzir palavras à sua raiz

Texto original: "Eu amo DS!"

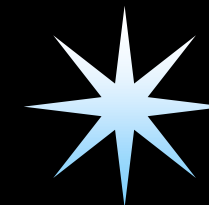
Após lowercasing: ["eu", "amo", "ds", "!"]

Lemmatization

Reduz a palavra à sua forma canônica

"Amasse" → Stemming: "am"

Lemmatization: "amar"



Cada etapa ajuda a preparar o texto para análise!

-
- # Texto Bruto (Original)
-

“Eu estou aprendendo processamento de linguagem natural!”

“Os algoritmos de machine learning são incríveis.”

“O cachorro correu rapidamente pelo parque.”

“Hoje eu comprei um livro sobre NLP.”

“As árvores estão balançando com o vento forte.”

-
- # Lowercasing
-

”eu estou aprendendo processamento de linguagem natural!”

”os algoritmos de machine learning são incríveis.”

”o cachorro correu rapidamente pelo parque.”

”hoje eu comprei um livro sobre nlp.”

”as árvores estão balançando com o vento forte.”

-
- # Remoção de Pontuação
-

”eu estou aprendendo processamento de linguagem natural”

”os algoritmos de machine learning são incríveis”

”o cachorro correu rapidamente pelo parque”

”hoje eu comprei um livro sobre nlp”

”as árvores estão balançando com o vento forte”

-
- # Tokenization
-

[eu, estou, aprendendo, processamento, de, linguagem, natural]

[os, algoritmos, de, machine, learning, sao, incríveis]

[o, cachorro, correu, rapidamente, pelo, parque]

[hoje, eu, comprei, um, livro, sobre, nlp]

[as, árvores, estão, balançando, com, o, vento, forte]

-
- # Stopwords
-

aprendendo processamento linguagem natural

algoritmos machine learning incríveis

cachorro correu rapidamente parque

comprei livro nlp

arvores balançando vento forte

-
- # Stemming
-

aprend, process, linguag, natur

algorithm, machin, learn, incriv

cachorr, corr, rapid, parqu

compr, livr, nlp

arvor, balanc, vent, fort

-
- # Lemmatization
-

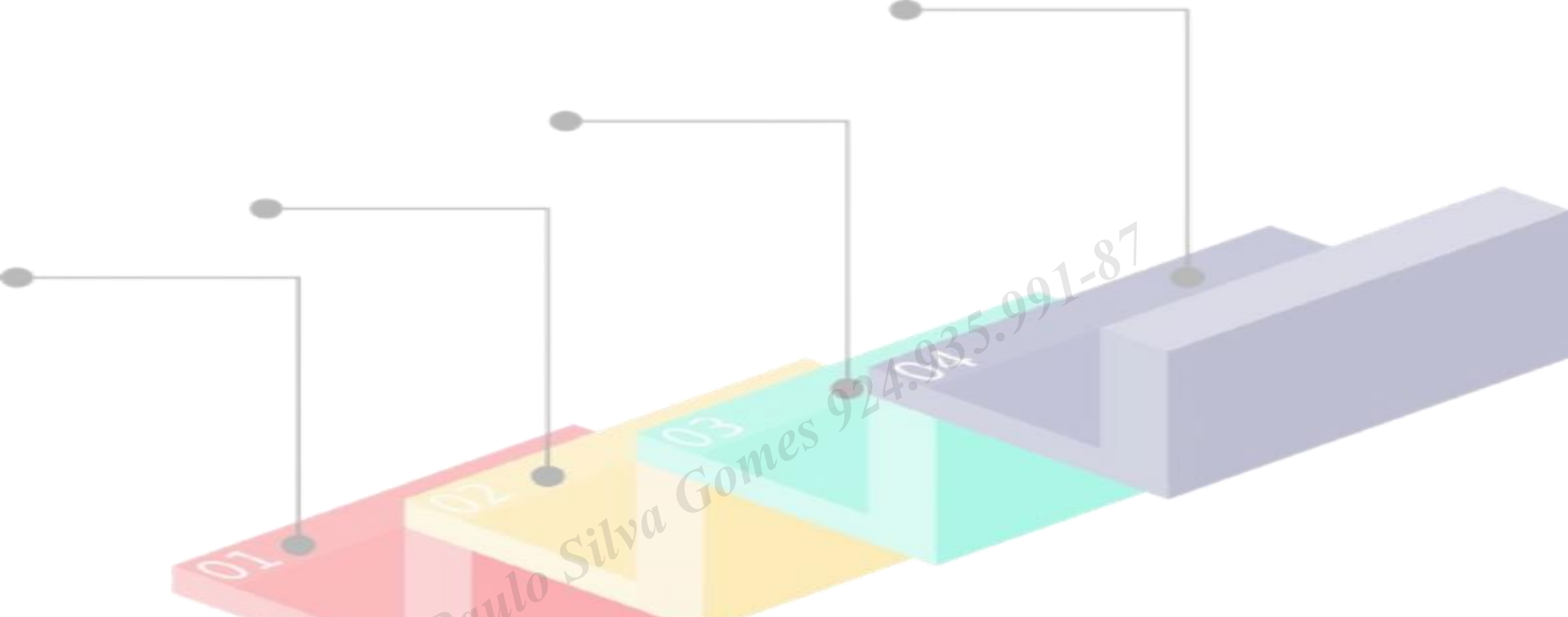
aprender, processamento, linguagem, natural

algoritmo, machine, learning, incrível

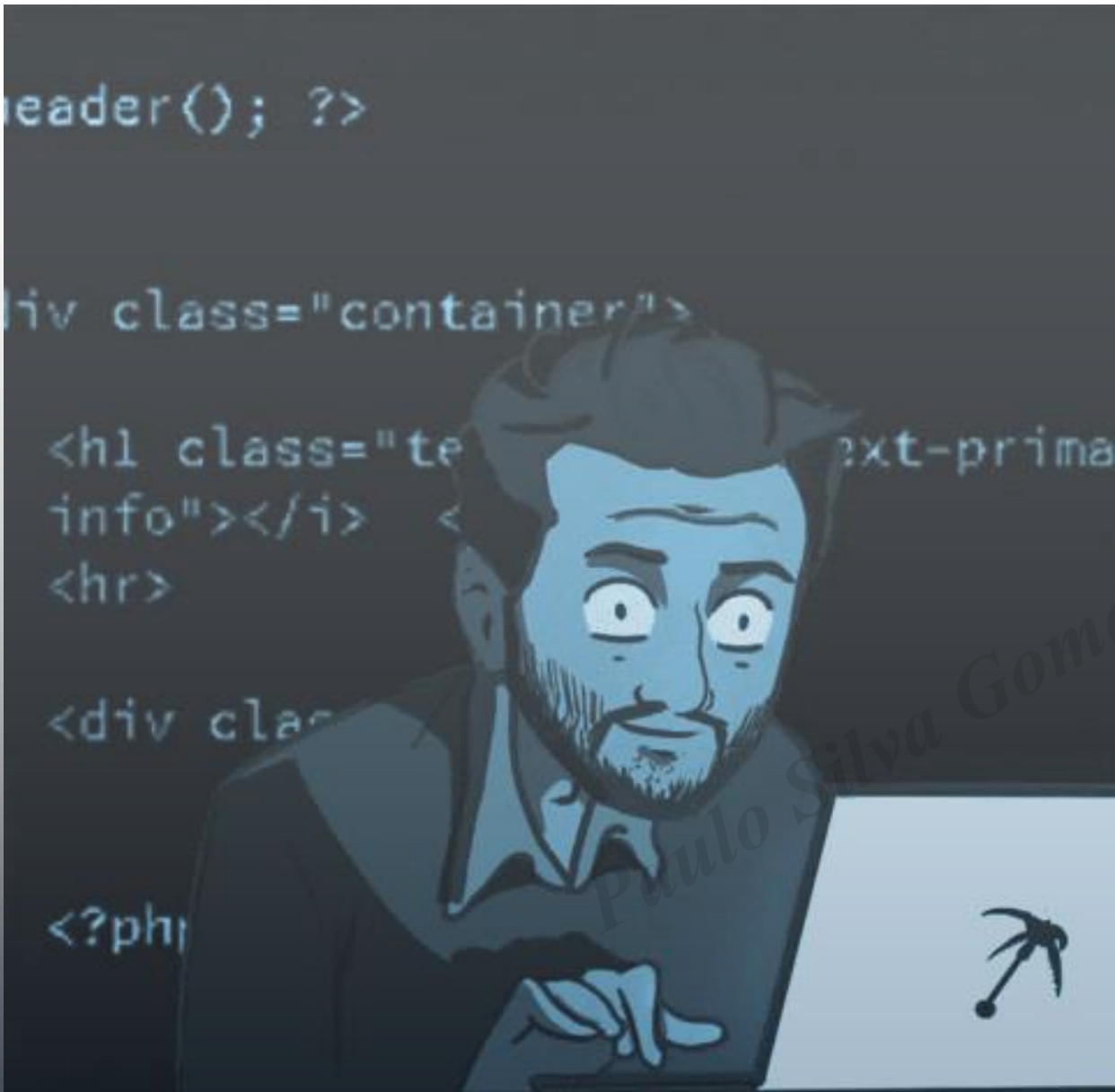
cachorro, correr, rápido, parque

comprar, livro, nlp

árvore, balançar, vento, forte



Cada etapa melhora a preparação dos dados para análise.



Vamos processar um texto

MBAUSP ESALQ

Obrigado!

[Professor Guilherme Lima | linkedin.com/in/guilherme-lima-developer](https://www.linkedin.com/in/guilherme-lima-developer)