

PROCESSAMENTO DE LINGUAGEM NATURAL

An abstract background pattern consisting of thin, light-colored lines and small circles, resembling a circuit board or a neural network diagram, set against a dark red background.

Classificação de textos

TÓPICOS

1. **Introdução**
2. **Classificação de textos**
3. **Exemplos e Aplicações**

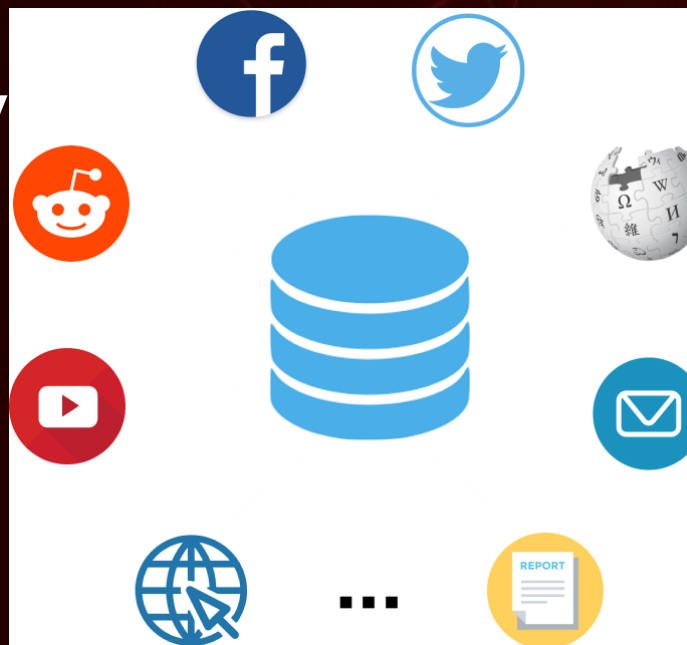


INTRODUÇÃO

- O conceito de *embeddings* refere-se a **uma representação em um espaço vetorial contínuo de menor dimensionalidade**, aprendida ou gerada a partir de uma representação de maior dimensionalidade
- As **embeddings** podem ser aprendidas/geradas para variáveis discretas (ex.: documentos, sentenças e palavras)
- Ao gerar uma representação de menor dimensionalidade, **deve-se tentar preservar as características do espaço de maior dimensionalidade**
- Ao gerar um espaço de menor dimensionalidade, o **processamento** por parte dos algoritmos de aprendizado de máquina torna-se **mais rápido e consome menos espaço**

INTRODUÇÃO

- A análise inteligente de textos tem muito destaque e muitas aplicações nos dias atuais
- Isso pelo fato de que dados textuais são comumente utilizados para transmitir/trocar informações:
 - *E-mails*
 - Artigos em jornais (notícias) e revistas
 - Relatórios e livros
 - Mensagens em aplicativos
 - Postagens em redes sociais
 - Páginas web



INTRODUÇÃO

- Além da facilidade de trocar ou armazenar informações textuais, o formato textual ocupa menos espaço que outros formatos, como o áudio, por exemplo



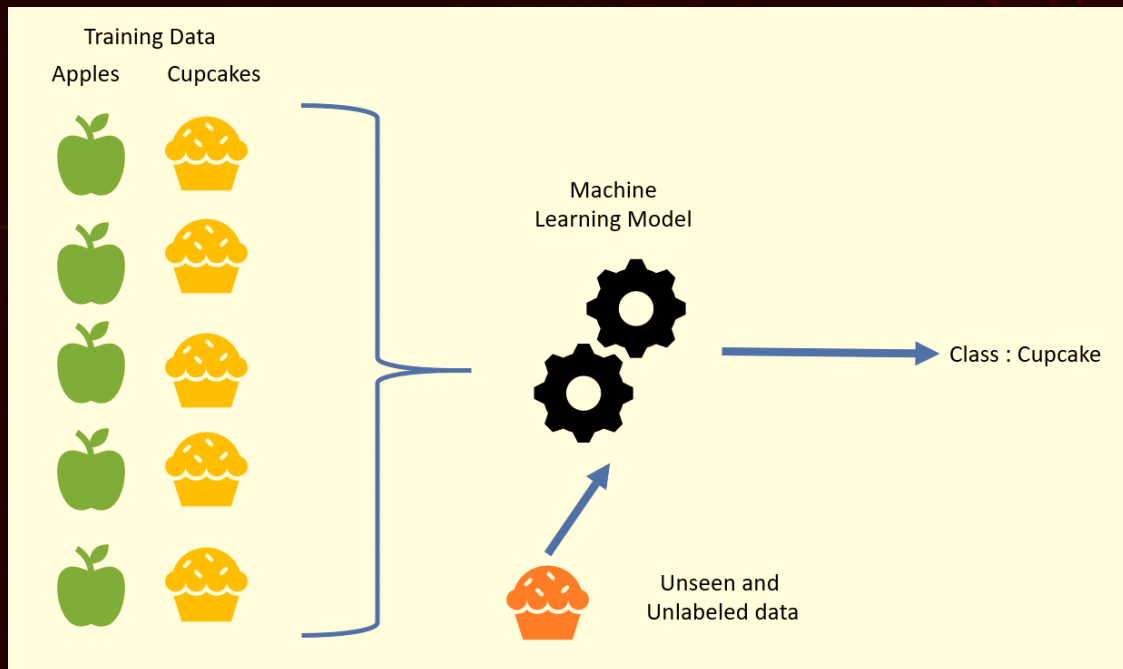
- A análise do conteúdo, tanto computacionalmente quanto por um humano, também é mais rápida
- Além disso, o formato textual permite consultar e recuperar informação mais facilmente

INTRODUÇÃO

- **Dar ao computador a habilidade de compreender conteúdos textuais pode ser útil para diversas aplicações, para o entendimento do comportamento humano, e para dar vantagem competitiva às empresas**
- **Exemplos de aplicações envolvendo a análise inteligente de textos:**
 - **Organização automática de documentos**
 - **Filtros de Spam ou o roteamento de e-mails para departamentos específicos**
 - **Análise de Sentimentos**
 - **Sumarização de documentos**
 - **Recomendação**
 - **Deteccção de intenções de suicídio ou determinar o perfil psicológico de acordo com as postagens**

CLASSIFICAÇÃO DE TEXTOS

- A **classificação automática de textos** é uma das tarefas mais utilizadas quando envolve textos **escritos em linguagem natural e inteligência artificial**, i.e., a análise inteligente de textos
- A classificação automática é a **atribuição de um rótulo** (valor discreto) a **um exemplo não rotulado** (ou novo exemplo)



CLASSIFICAÇÃO DE TEXTOS

- Portanto, a **classificação automática de textos** envolve atribuir automaticamente um rótulo a um texto completo (documento), ou ainda, a partes do texto (parágrafos, sentenças ou sequências de palavras)



- A classificação automática é **bem flexível** no sentido de que para um mesmo texto, diferentes rótulos podem ser atribuídos, de acordo com o gosto do freguês

EXEMPLOS

Rótulos pré-definidos

Paulista

Carioca

Gaúcho

Brasileiro

...

Esportes Futebol

São Paulo lista revelação da base para jogo deste sábado

O ESTADO DE S. PAULO

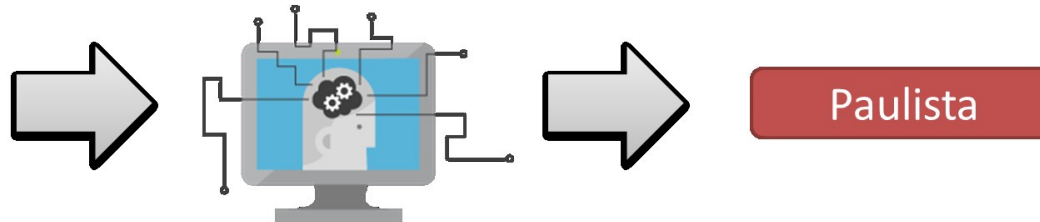
04 Março 2016 | 18h 28 - Atualizado: 04 Março 2016 | 18h 28

Lucas Fernandes, de 18 anos, ganha chance pela primeira vez

A lista de relacionados do [São Paulo](#) para o jogo deste sábado, contra o São Bernardo, tem uma novidade. O meia Lucas Fernandes, de 18 anos, acaba de ser promovido das categorias de base e recebeu a chance do técnico Edgardo Bauza para ficar ao menos no banco de reservas para a partida que será no Pacaembu, pelo [Campeonato Paulista](#).

Fonte: www.estadao.com.br

Classificador Automático



EXEMPLOS

Rótulos pré-definidos

São Paulo

Palmeiras

Corinthians

Santos

...

Esportes Futebol



A+
A-



São Paulo lista revelação da base para jogo deste sábado

O ESTADO DE S. PAULO

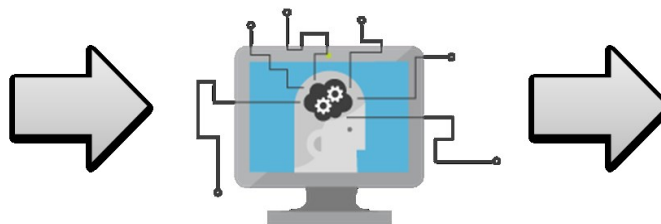
04 Março 2016 | 18h 28 - Atualizado: 04 Março 2016 | 18h 28

Lucas Fernandes, de 18 anos, ganha chance pela primeira vez

A lista de relacionados do [São Paulo](#) para o jogo deste sábado, contra o São Bernardo, tem uma novidade. O meia Lucas Fernandes, de 18 anos, acaba de ser promovido das categorias de base e recebeu a chance do técnico Edgardo Bauza para ficar ao menos no banco de reservas para a partida que será no Pacaembu, pelo [Campeonato Paulista](#).

Fonte: www.estadão.com.br

Classificador Automático



São Paulo

EXEMPLOS

Rótulos pré-definidos

Interessante

Não Interessante

Esportes Futebol



A+
A-



São Paulo lista revelação da base para jogo deste sábado

O ESTADO DE S. PAULO

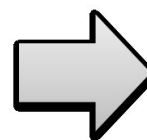
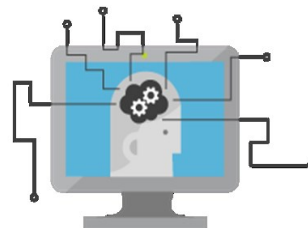
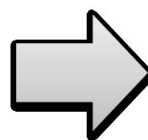
04 Março 2016 | 18h 28 - Atualizado: 04 Março 2016 | 18h 28

Lucas Fernandes, de 18 anos, ganha chance pela primeira vez

A lista de relacionados do [São Paulo](#) para o jogo deste sábado, contra o São Bernardo, tem uma novidade. O meia Lucas Fernandes, de 18 anos, acaba de ser promovido das categorias de base e recebeu a chance do técnico Edgardo Bauza para ficar ao menos no banco de reservas para a partida que será no Pacaembu, pelo [Campeonato Paulista](#).

Fonte: www.estadão.com.br

Classificador Automático



Interessante

EXEMPLOS

Rótulos pré-definidos

Melhor Time do Mundo

Resto

Esportes Futebol



A+
A-



São Paulo lista revelação da base para jogo deste sábado

O ESTADO DE S. PAULO

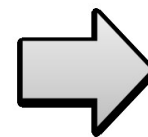
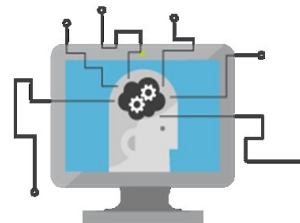
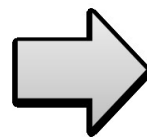
04 Março 2016 | 18h 28 - Atualizado: 04 Março 2016 | 18h 28

Lucas Fernandes, de 18 anos, ganha chance pela primeira vez

A lista de relacionados do [São Paulo](#) para o jogo deste sábado, contra o São Bernardo, tem uma novidade. O meia Lucas Fernandes, de 18 anos, acaba de ser promovido das categorias de base e recebeu a chance do técnico Edgardo Bauza para ficar ao menos no banco de reservas para a partida que será no Pacaembu, pelo [Campeonato Paulista](#).


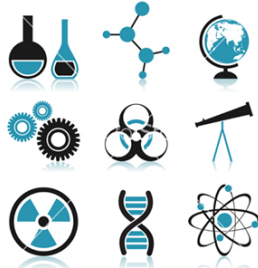

Fonte: www.estadao.com.br

Classificador Automático



Melhor Time do
Mundo

OUTROS EXEMPLOS

DOMÍNIO	RÓTULOS	
	<div>Spam</div> <div>Não Spam</div>	<div>Finanças</div> <div>SAC</div> <div>RH</div> <div>...</div>
	<div>Geografia</div> <div>Biologia</div> <div>Computação</div> <div>...</div>	<div>Banco de Dados</div> <div>Sistemas Distribuídos</div> <div>Inteligência Artificial</div> <div>...</div>
	<div>Negativo</div> <div>Neutro</div> <div>Positivo</div>	<div>Ofensivo</div> <div>Não Ofensivo</div>

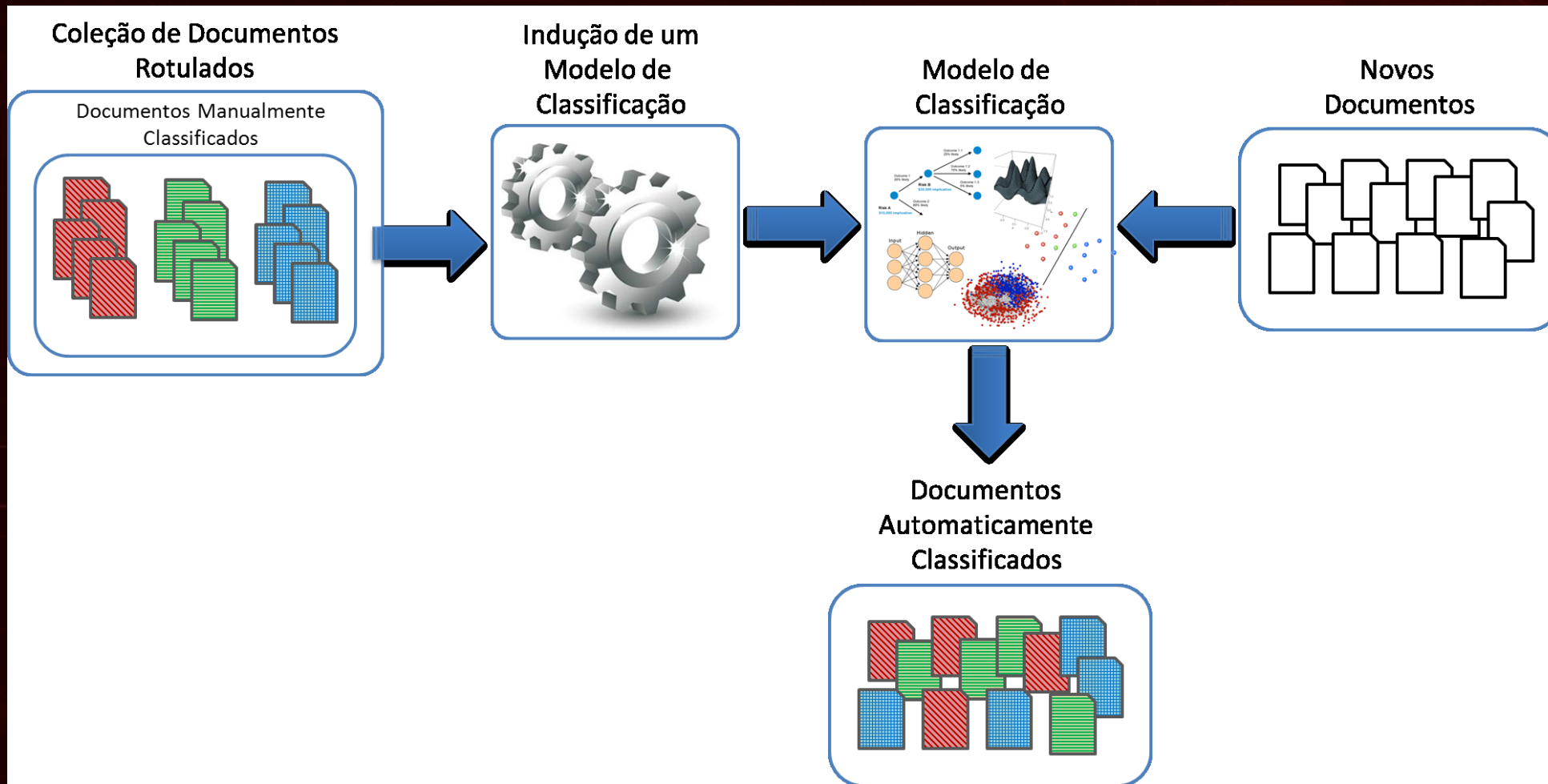
CLASSIFICAÇÃO DE TEXTOS

- Devido ao grande volume de dados gerado nos dias atuais, é muitas vezes humanamente impossível **organizar, analisar e extrair conhecimento** embutido nas informações textuais
- **A classificação automática de textos pode apoiar em tais tarefas**
- **Permite organizar e gerenciar**
 - Roteamentos, Filtragens e Atribuições de metadados
- **Permite analisar e extrair conhecimento**
 - Estatísticas dos rótulos atribuídos aos documentos
 - Análises descritivas e preditivas de *WebSensors*
 - Conjunto de regras ou valores que sumarizam os padrões presentes na coleção de documentos

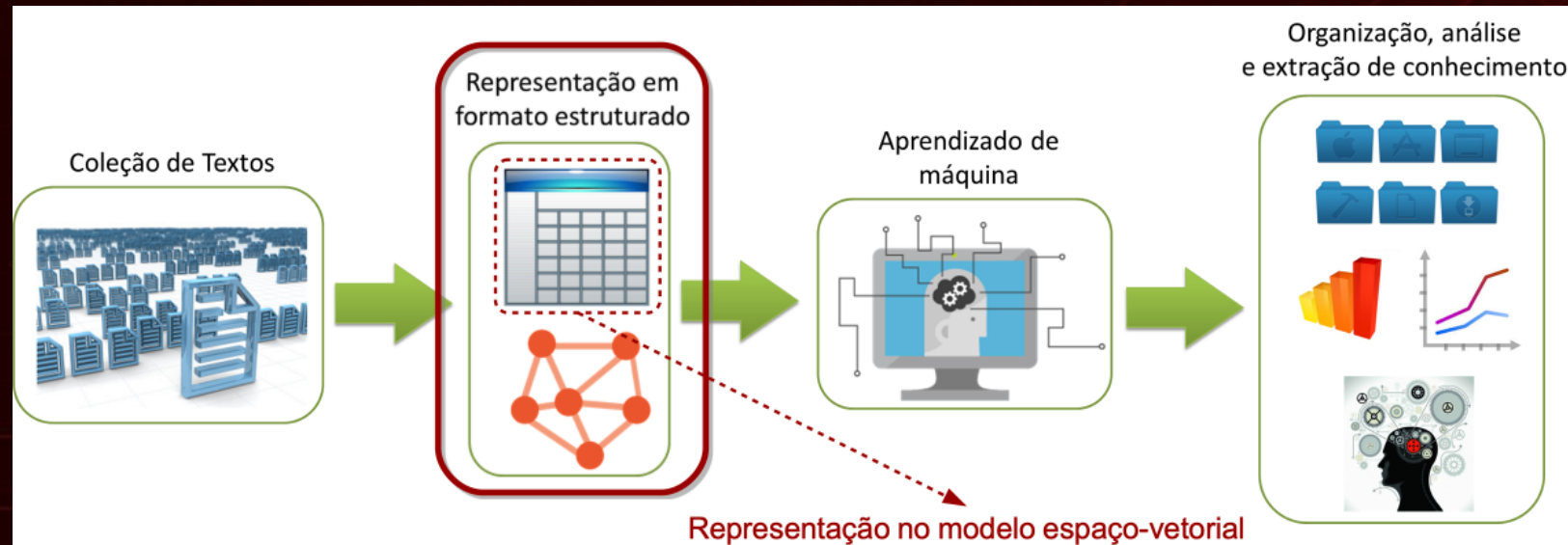
EXEMPLOS DE APLICAÇÃO

- Organização e recuperação de documentos
- Filtragem e organização de notícias
- Recuperação de informação
- Sistemas de Recomendação
- Classificação de e-mails, detecção de *spams* e resposta automática
- Mineração de opinião/Análise de sentimentos
- Identificação de crimes
- Respostas às questões
- Classificação de intenções
- Aplicações em outros domínios com características semelhantes às textuais

CLASSIFICAÇÃO TEXTUAL COM AM

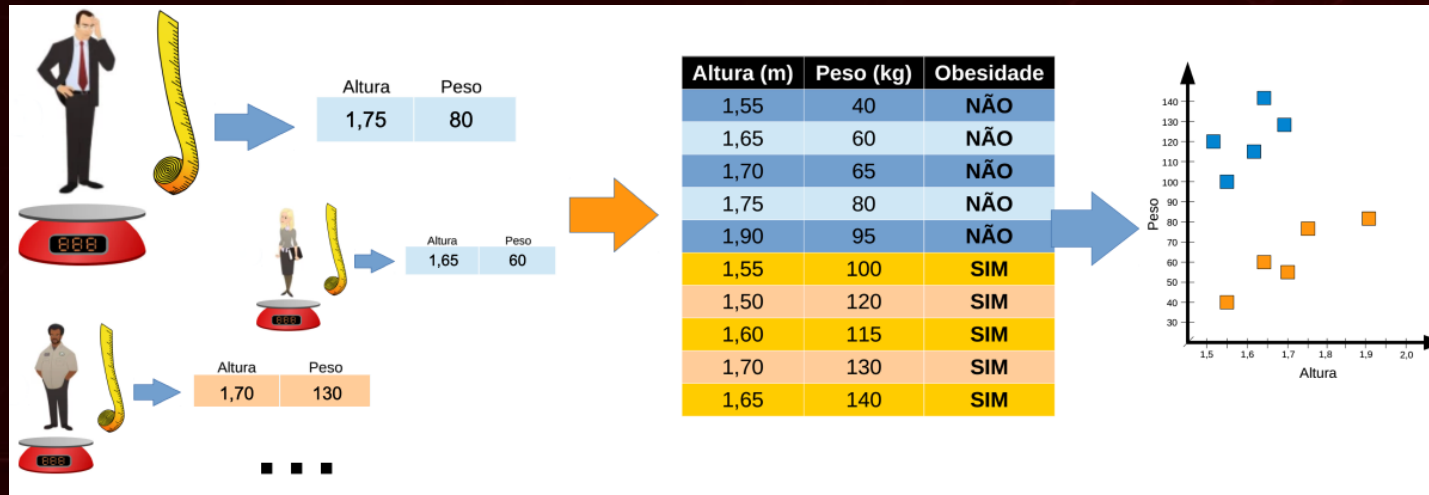


ABORDAGEM CLÁSSICA



- Representações (Formato estruturado)
- Limpeza
- Padronização dos termos

CLASSIFICAÇÃO TEXTUAL COM AM

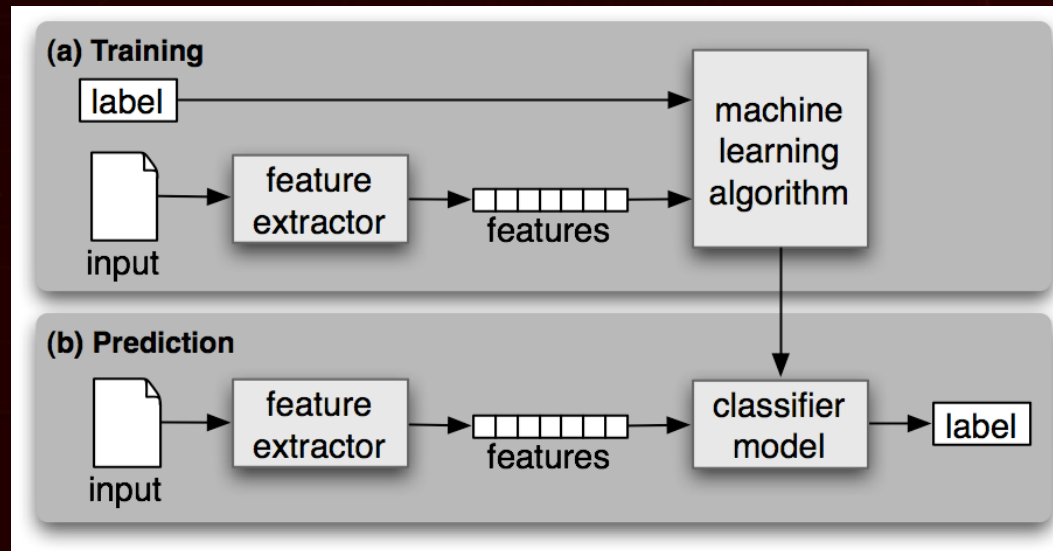


Texto		Classe	
Consumer_complaint_narrative		Product	
0	Diversified Consultants have acquired a debt L...	Debt collection	
1	We refinanced a mortgage that was help by PNC ...	Mortgage	
2	I have tried to open a checking account with T...	Credit reporting, credit repair services, or o...	
3	Transunion has not been compliant with my reso...	Credit reporting	
4	Hello : Here is a concise summary of what has ...	Credit reporting, credit repair services, or o...	
...	
45202	I have a public record & Child Support entry L...	Credit reporting, credit repair services, or o...	
45203	My home mortgage closing is pending. Wells Fai...	Mortgage	
45204	Due Diligence from Equifax was not performed a...	Credit reporting	
45205	I had bought a 1998 XXXX from a buy here pay h...	Consumer Loan	
45206	I have an autopay set up to draw the amount eq...	Credit card or prepaid card	


?

	Atr. 1?	Atr. 2?	Atr. 3?	...	Atr. N-1?	Atr. N?	Classe
0	?	?	?	...	?	?	Debt collection
1	?	?	?	...	?	?	Mortgage
2	?	?	?	...	?	?	Credit reporting, ...
3	?	?	?	...	?	?	Credit reporting
4	?	?	?	...	?	?	Credit reporting, ...
...	?	?	?	...	?	?	...
45202	?	?	?	...	?	?	Credit reporting, ...
45203	?	?	?	...	?	?	Mortgage
45204	?	?	?	...	?	?	Credit reporting
45205	?	?	?	...	?	?	Consumer Loan
45206	?	?	?	...	?	?	Credit card or ...

CLASSIFICAÇÃO TEXTUAL COM AM

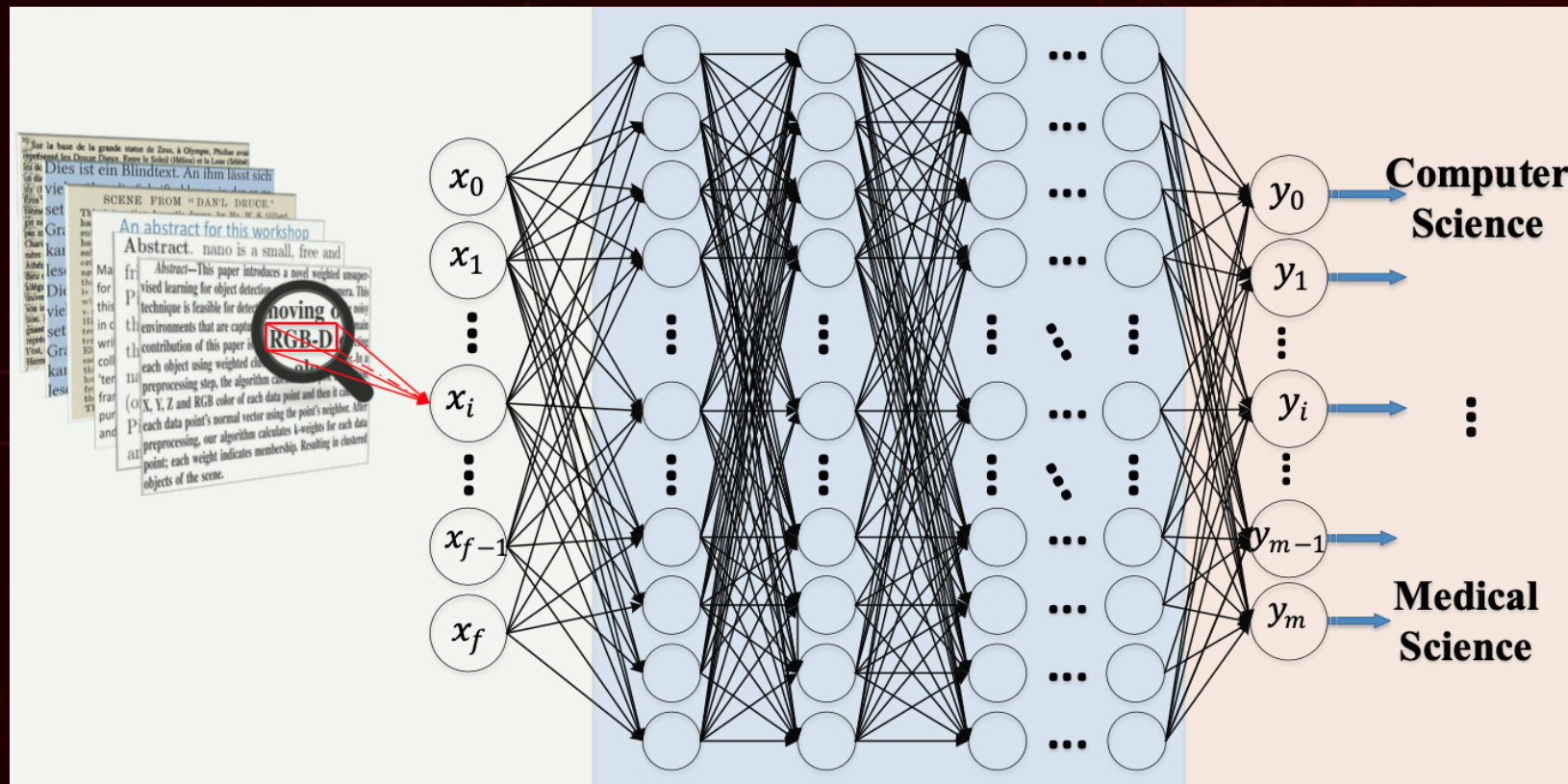


Texto			Classe	
	Consumer_complaint_narrative	Product		
0	Diversified Consultants have acquired a debt L...	Debt collection		
1	We refinanced a mortgage that was help by PNC ...	Mortgage		
2	I have tried to open a checking account with T...	Credit reporting, credit repair services, or o...		
3	Transunion has not been compliant with my rese...	Credit reporting		
4	Hello : Here is a concise summary of what has ...	Credit reporting, credit repair services, or o...		
...		
45202	I have a public record & Child Support entry L...	Credit reporting, credit repair services, or o...		
45203	My home mortgage closing is pending. Wells Fai...	Mortgage		
45204	Due Diligence from Equifax was not performed a...	Credit reporting		
45205	I had bought a 1998 XXXX from a buy here pay h...	Consumer Loan		
45206	I have an autopay set up to draw the amount eq...	Credit card or prepaid card		



	Atr. 1?	Atr. 2?	Atr. 3?	...	Atr. N-1?	Atr. N?	Classe
0	?	?	?	...	?	?	Debt collection
1	?	?	?	...	?	?	Mortgage
2	?	?	?	...	?	?	Credit reporting, ...
3	?	?	?	...	?	?	Credit reporting
4	?	?	?	...	?	?	Credit reporting, ...
...	?	?	?	...	?	?	...
45202	?	?	?	...	?	?	Credit reporting, ...
45203	?	?	?	...	?	?	Mortgage
45204	?	?	?	...	?	?	Credit reporting
45205	?	?	?	...	?	?	Consumer Loan
45206	?	?	?	...	?	?	Credit card or ...

ABORDAGEM ATUAL



O QUE VIMOS?

- **Introdução**
- **Classificação de textos**
- **Exemplos e Aplicações**



PRÓXIMA VIDEOAULA

- **Redes Neurais Artificiais**

