

Data science aplicada aos negócios

Prof. Dr. Álvaro Campos
Ferreira

alvaro.ferreira@idp.edu.br

Aprendizado de Máquina

Aprendizado de Máquina



Aprendizado de Máquina é a arte e a ciência de programar computadores para que eles possam aprender com dados.

Aprendizado de Máquina

Aprendizado de Máquina é o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados.

Arthur Samuel, 1959.

Aprendizado de Máquina

É dito que um programa de computador aprende com a Experiência E com relação a alguma Tarefa T e uma medida de Performance P , se sua performance em T , como medido por P , melhora com a experiência E .

Tom Mitchell, 1997.

Aprendizado de Máquina

O filtro de Spam é um programa de Aprendizado de Máquina que utiliza exemplos de e-mails rotulados como Spam para marcar e-mails como Spam:

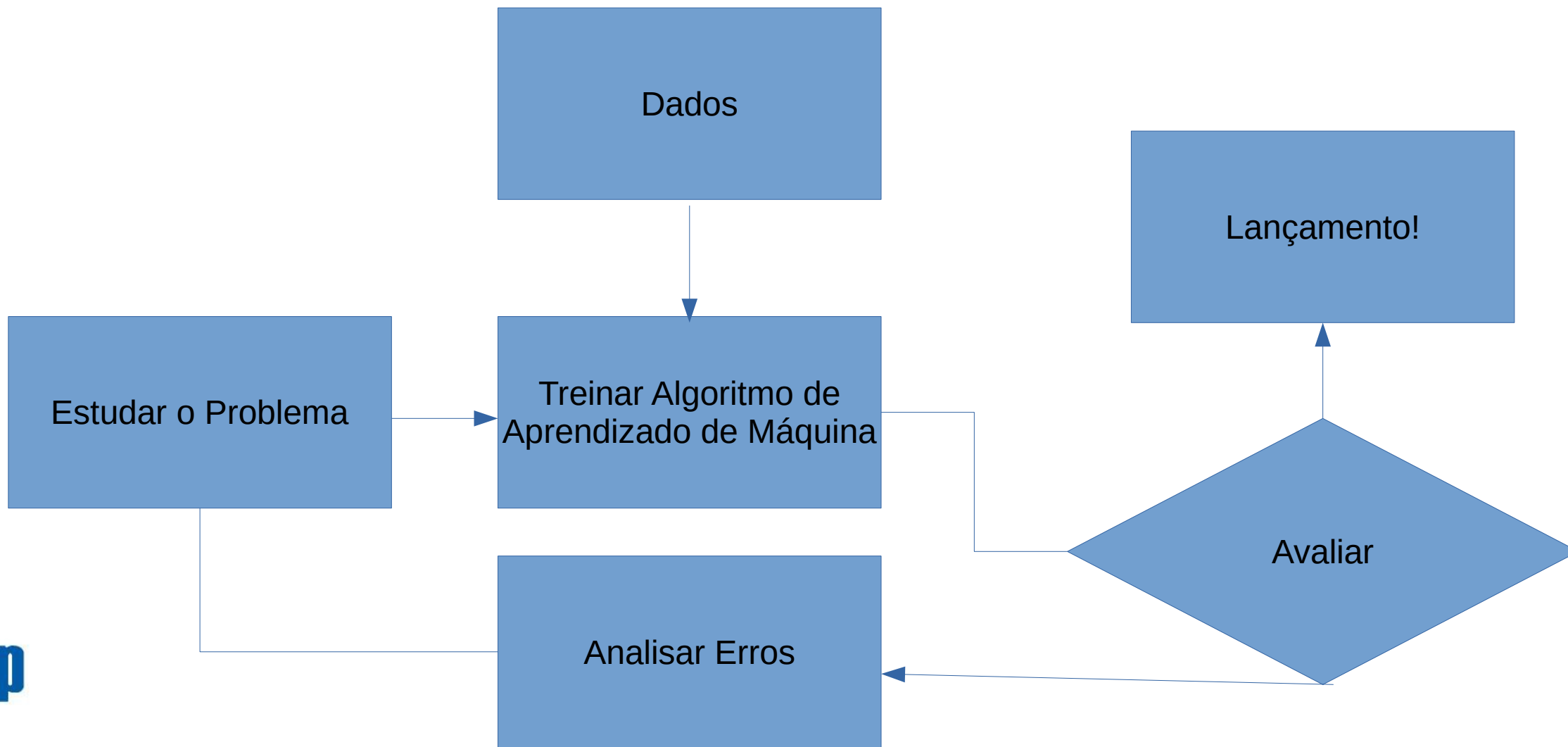
- Tarefa T: Marcar Spam
- Experiência E: Dados de treinamento (e-mails)
- Performance P: Precisa ser definida

Aprendizado de Máquina

A Performance P do sistema de Aprendizado de Máquina deve ser definida para o caso estudado. Por exemplo, podemos usar a relação de e-mails classificados corretamente.

- Essa medida é chamada precisão

A abordagem do Aprendizado de Máquina



A abordagem do Aprendizado de Máquina

Em resumo, Aprendizado de Máquina é ótimo para:

- Problemas para os quais soluções tradicionais exigem muitos ajustes manuais ou longas listas de regras
- Problemas complexos para os quais não existem soluções adequadas com a abordagem tradicional.

A abordagem do Aprendizado de Máquina

Em resumo, Aprendizado de Máquina é ótimo para:

- Ambientes com grande flutuação, que exigem adaptação do algoritmo
- Obter insights sobre problemas complexos e grandes quantidades de dados

Fluxo de Trabalho para Aprendizado de Máquina

Fluxo de Trabalho para Aprendizado de Máquina

Os passos para realizar um projeto de Aprendizado de Máquina:

1. Investigar o quadro geral
2. Obter os dados.
3. Descobrir e visualizar os dados para obter insights.

Fluxo de Trabalho para Aprendizado de Máquina

4. Preparar os dados para algoritmos de Aprendizado de Máquina.
5. Selecionar um modelo e treina-lo.
6. Ajustar o modelo.
7. Apresentar a sua solução.
8. Lançar, monitorar e manter seu sistema.

Tipos de Sistemas de Aprendizizado de Máquina

Tipos de Sistemas de Aprendizado de Máquina

Vamos classificar em categorias baseadas em se:

- O sistema é treinado com supervisão humana
- Pode aprender incrementalmente a partir de um fluxo de dados
- Funciona a partir da comparação simples com valores conhecidos ou a partir de um modelo

Aprendizado Supervisionado

Aprendizado Supervisionado

Em aprendizado supervisionado, os dados de treinamento possuem as soluções desejadas, chamadas rótulos (labels).

- Por exemplo: o conjunto de dados contém e-mails Spam e normais, marcados de acordo
Essa é uma tarefa de **Classificação**.

Aprendizado Supervisionado

Outra tarefa típica do aprendizado supervisionado é a **Regressão**.

Aprendizado Supervisionado

Regressão é a tarefa de prever um valor numérico alvo, como o preço de um carro, a partir de um conjunto de características chamadas preditores.

- Para treinar o sistema, você precisa de muitos exemplos de carros, incluindo seus preditores e rótulos (preços)

Aprendizado Supervisionado

Note que alguns algoritmos de regressão podem ser utilizados para classificação e vice versa.

- Por exemplo, Regressão Logística é utilizada comumente para classificação pois seu output é a probabilidade de um elemento pertencer a uma classe

Aprendizado Não Supervisionado

Aprendizado Não Supervisionado

Em aprendizado não supervisionado, o conjunto de dados não está rotulado (unlabeled).

- Por exemplo: Você tem muitos dados sobre os visitantes do seu Blog. Talvez você queira agrupar seus usuários por interesse para realizar marketing direcionado

Aprendizado Não Supervisionado

Detecção de anomalia é outra tarefa de aprendizado não supervisionado em que o sistema é treinado com instâncias normais e quando ele vê uma nova instância, ele sabe dizer se é uma anomalia.

- Detecção de fraude

Aprendizado Não Supervisionado

Regras de Associação é quando o objetivo é descobrir relações interessantes entre os atributos.

- Por exemplo, pessoas que compram feijão preto e louro também tendem a comprar cebola, então faz sentido colocar esses itens próximos uns dos outros.

Aprendizado por Reforço

Aprendizado por Reforço

Em aprendizado por reforço, o sistema é chamado *agente* e ele pode observar o *ambiente*, selecionar e realizar ações para receber *recompensas*.

- Deve aprender sozinho a melhor estratégia, chamada *política*, para receber mais recompensas ao longo do tempo.

Aprendizado por Reforço

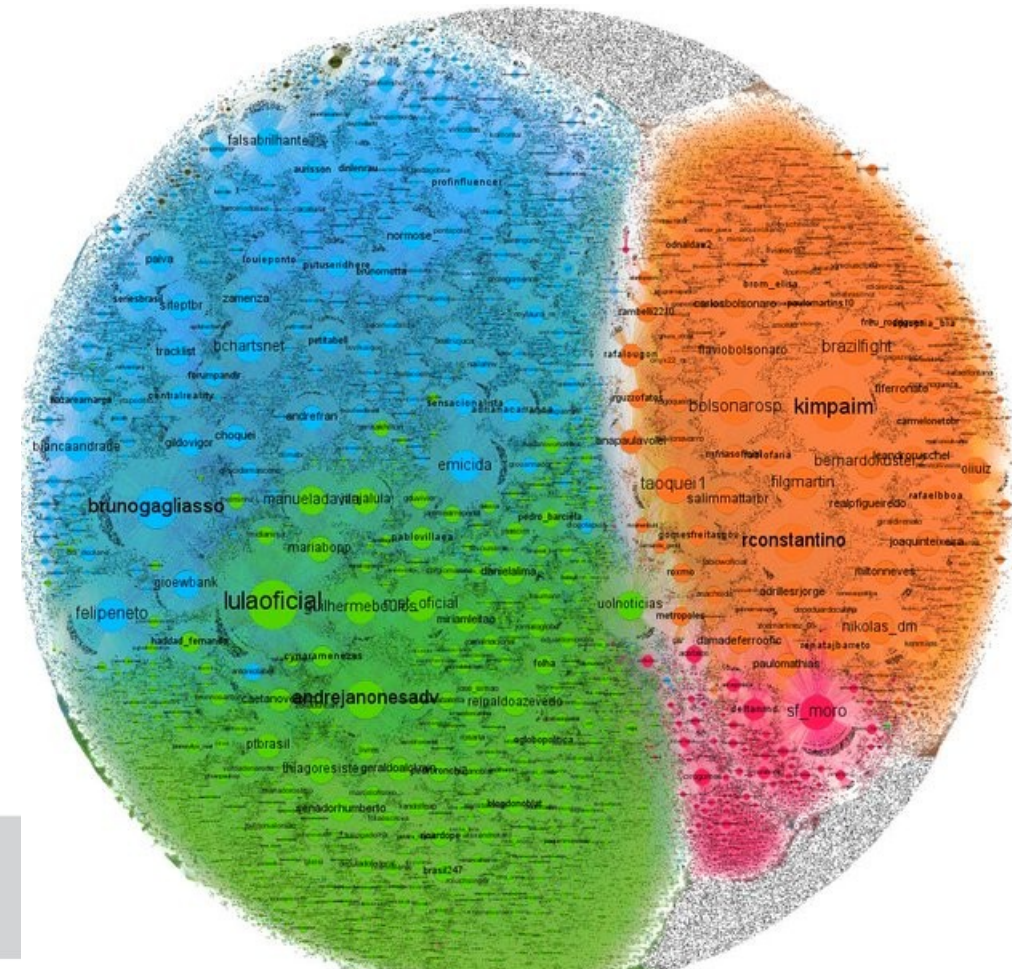
Políticas definem que ação o agente deve escolher quando está em uma dada situação.

- AlphaGo
- Marl/O

<https://www.youtube.com/watch?v=qv6UVOQ0F44>

Visualizações

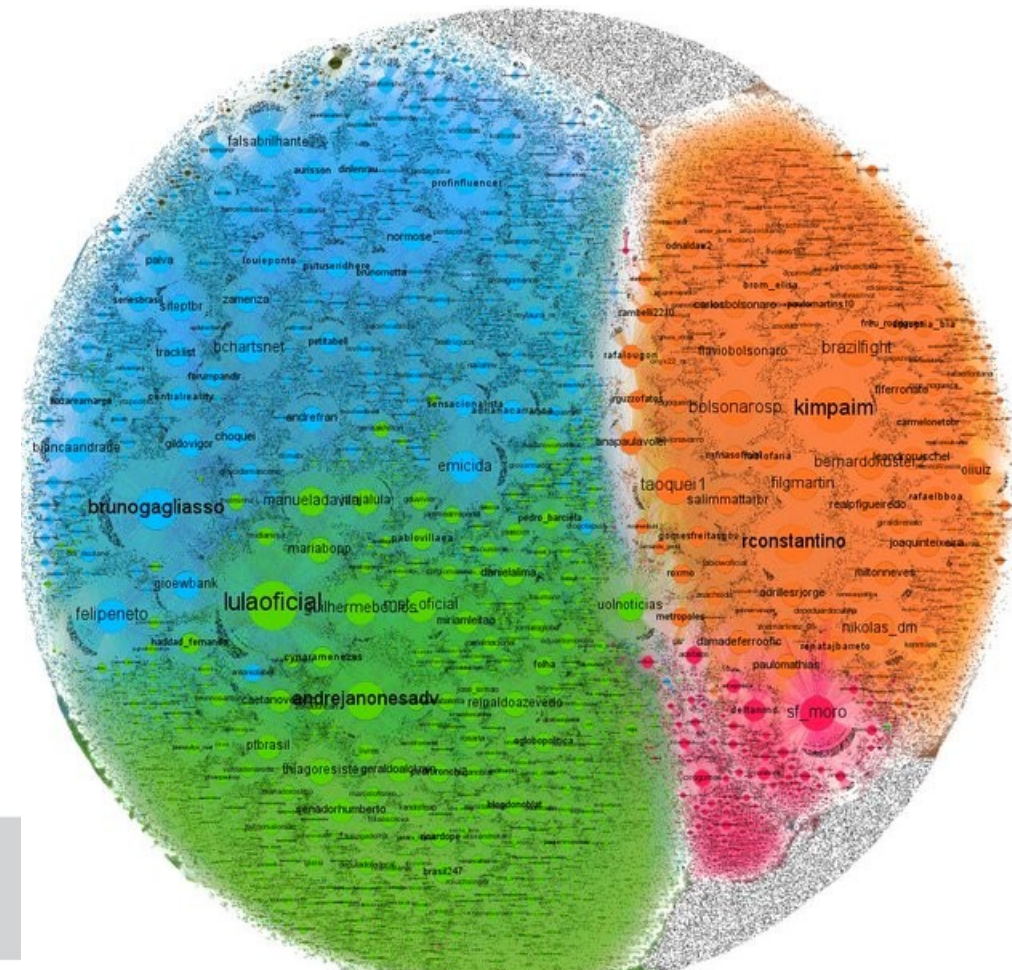
Visualizações



Análise de interações no Twitter usando grafos.

- Fabio Malini (@fabiomalini) “Só hoje, por curiosidade, processei os 3.404.286 tuítes gerados sobre o Lula durante e após a entrevista no Jornal Nacional.”

Visualizações



“E visualizei os agrupamentos, representados pelas cores no grafo. Um dado me chama atenção: 40% dos Rts vêm dos azuis (antibolsonaros).”



INSTITUTO BRASILEIRO DE ENSINO,
DESENVOLVIMENTO E PESQUISA