A principal coisa que Machine Learning (Aprendizado de máquina) faz é encontrar um padrão de dados. Em seguida usa esses padrões para prever o futuro.

Exemplos de uso do Machine Learning:

- Usar para detectar fraudes de cartão de crédito;

- Se o cliente mudará para um concorrente;

- Qual a hora de fazer manutenção preventiva em um robô de fábrica;

O aprendizado de máquina:

- Começa com dados que contêm padrões;

- Alimenta os dados em um algoritmo de aprendizado, que gera um modelo;

- Um modelo é funcionalidade, tipicamente código, que é capaz de reconhecer padrões quando apresentado com novos dados;

- As aplicações podem usar esse modelo, fornecendo novos dados, para ver se esses dados correspondem a padrões conhecidos

Quem está interessado em um Machine Learning?

Possui 3 grupos de pessoas interessados:

- Líderes empresariais;

- Desenvolvedores de software;

- Cientista de dados (Estatísticas, Software de Machine Learning, domínio de problemas)

Pontos principais:

- Machine Learning permite encontrar padrões nos dados existentes e criar e usar um modelo que reconheça esses padrões nos novos dados.

- Machine Learning se tornou popular.

- Machine Learning provavelmente pode ajudar a organização

O primeiro PROBLEMA no Machine Learning é decidir qual pergunta fazer, o processo mais importante é decidir qual pergunta fazer.

Analisando o processo de Machine Learning com mais detalhes:

**Primeiro** escolhe os dados que deseja trabalhar.

Mas quase sempre não estão na forma correto, possui duplicidade, falta de dados, tem material extra. Normalmente precisa aplicar um **pré-processamento** nos dados e os produtos do Machine Learning fornecem uma variedade de módulos de pré-processamento.

Os resultados são dados que foram trabalhados. Depois de ter os dados é aplicado algoritmos de aprendizado de máquina e novamente os produtos do Machine Learning fornecem algoritmos de aprendizado de máquina e o resultado é um modelo.

Esses processos são feitos até gerar um modelo que atenda a necessidade e depois disso as aplicações podem utilizar esse modelo.

**Terminologia**

Treinamento de dados: são os dados preparados para criar um modelo. Criar um modelo é chamado de Training Data (Treinamento de dados)

Existem duas grandes categorias do Machine Learning:

1ª Supervised Learning (Aprendizado Supervisionado): o valor que deseja prever está realmente nos dados de treinamento.

2ª Unsupervised Learning (Aprendizado não Supervisionado): o valor que deseja prever NÃO está nos dados de treinamento, os dados não estão marcados.

Ambas são usadas, mas a mais comum é a Supervised Learning.

**Pré-processamento de Dados com Supervised Learning**

Os dados podem ser Relacionais, Banco de dados NoSQL, Dados binários. Independentemente de onde venha, é necessário ler os dados com um módulo de pré-processamento de dados.

**Categorizando Problemas de Machine Learning:**

**Regressão**

**Classificação**

**Clustering (Agrupamento)**