

Machine learning é um ramo da inteligência artificial no qual pressupõe que uma máquina pode aprender com dados, e tomar decisões a partir de padrões, com o mínimo de Intervenção humana. Lembrando que aprender significa conseguir fazer melhor se algo parecido acontecer.

Há 3 tipos de aprendizagem: aprendizado supervisionado, aprendizado não supervisionado, e aprendizado por reforço. O aprendizado supervisionado é positivo pois o sistema aprende com respostas corretas previamente selecionadas por outra pessoa, o “supervisor”, esse método é bastante utilizado para realizar classificações. No método não supervisionado o sistema reconhece os padrões sem nenhum feedback prévio. Por último temos o aprendizado por reforço, no qual o sistema recebe punições ou recompensas de acordo com a resposta.

Apesar de ser mais fácil, não recomenda-se aprender por memorização, pois mesmo sendo mais fácil para as máquinas depois de um tempo se tornaria inviável em questão de armazenamento de dados. O ideal é aprender por generalização igual os humanos, pois assim não se necessita de elevados custos para o armazenamento de dados, no entanto para as máquinas isto é mais difícil pois a diversas maneiras de generalizar.

Existem diversos algoritmos feitos para cada tipo de aprendizagem, no aprendizado supervisionado existem: Árvore de decisão, K-Nearest Neighbor (KNN), Support Vector Machines (SVM) e muitos outros.

Para aprendizado supervisionado há também redes neurais, que visa imitar o cérebro humano, pois como dito antes o cérebro humano é muito bom em reconhecer padrões, ao contrário das máquinas, que são ótimas com cálculos, a velocidade com que a rede neural aprende depende de vários fatores, entre eles estão a complexidade da rede, o tamanho, a arquitetura, o algoritmo e a precisão desejada.

No aprendizado não supervisionado são comuns algoritmos de agrupamento, como clusterização, Basic Sequential Algorithmic Scheme (BSAS), clusterização hierárquica e o algoritmo K-Means que é um algoritmo baseado em otimização.

A carreira em engenharia de dados é bem ampla, e quem ingressa nela pode ser cientista de dados, engenheiro de dados, analista de dados, arquiteto de dados, analista de negócios, administrador de banco de dados, estatístico, entre outros.

Para trabalhar com mais eficiência há a ferramenta Knime, que permite que o usuário faça a análise de dados, relatório e integração de dados.