

## Podcast

**Disciplina:** Linguagens de programação para ciência de dados

**Título do tema:** Machine Learning em Python

**Autoria:** Yuri Vasconcelos de Almeida Sá

**Leitura crítica:** Henrique Salustiano Silva

### Abertura:

Olá! No podcast de hoje vamos falar sobre Machine Learning, Estatística e as Modas que tem por aí.

A internet gerou muito Buzz nos últimos anos em relação ao termo Machine Learning. Como se fosse algo extraordinário ou até mesmo assustador. Se pegarmos o termo em sua definição, machine learning nada mais é do que fazer a máquina aprender com os dados apresentados a ela, e aqui começam os primeiros equívocos. Pode parecer moderno, mas existe machine learning desde que os computadores começaram a fazer cálculos estatísticos. Até mesmo uma simples regressão linear de duas variáveis pode ser considerada machine learning.

Então podemos dizer que o machine learning existe desde que os computadores foram criados, uma história longa até.

Mas é claro que esse medo do machine learning tem fundamento sim. O que acontece é que com o avanço da computação e dos modelos estatísticos, novas bibliotecas e métodos foram criados para explorar estes modelos criando verdadeiros monstros que avaliam quantidades monstruosas de dados com diversas fontes e concatenam tudo em um só resultado, como forma de decisão. O medo mora na incerteza, e alguns destes modelos são chamados de caixa-preta, ou seja, seu funcionamento não pode ser descrito, a formação do resultado não pode ser examinada, e então mora no nosso imaginário a revolução das máquinas, dos exterminadores e etc... Afinal, nós fomos bombardeados com estes folclores a vida inteira.

Esta evolução impulsionada pelo barateamento do hardware criou outras formas de pensar nos sistemas computacionais de aprendizado. Existe hoje o Deep Learning, que seria o aprendizado profundo, utilizando muitos dados e treinando o computador para que faça sentido com estes dados, automatizando também o processo de escolha dos modelos estatísticos e seu encadeamento.

Este tipo de técnica está ficando cada vez mais comum e acessível, mas a compreensão do modo de aprender do computador só fica mais nebuloso e incompreensível, como estes modelos estão cada vez mais presentes em nossas vidas, ficamos com medo das decisões que possam influenciar nossas vidas.

Existe um forte movimento dentro dos círculos da ciência de dados para explicar as decisões e aumentar a acessibilidade ao funcionamento dos modelos de Machine Learning.

Este fino equilíbrio terá que ser eternamente ajustado, com a evolução do hardware e do software de aprendizado, mais e mais modelos e bibliotecas Caixa Preta aparecerão e novamente o esforço de tornar o método claro terá que ser feito.

De outro lado, os métodos mais modernos como Redes Neurais Convolucionais produzem resultados fabulosos com pouquíssimo código, porém são extremamente difíceis de serem explicadas, com múltiplas fases de processamento e um treinamento intenso, a técnica pode assustar quando erra em produção pois não saberíamos o que pode ter ocasionado o erro.

Uma vantagem é que boa parte destas bibliotecas tem seu código publicado sob licenças *open source*, onde podemos ver o código e criar nós mesmos saídas para tentar explicar melhor os processos que levam ao resultado.

Uma boa prática em geral é verificar se o modelo que você escolheu tem uma boa explicação sobre seu funcionamento!

### **Fechamento:**

Este foi nosso podcast de hoje! Até a próxima!