

# Tendências em *Machine Learning* *AutoML e MLOps*

Prof. Sandro Jerônimo

Fontes das Imagens: [www.shutterstock.com](http://www.shutterstock.com)  
ou próprio autor

PUC Minas Virtual



# *Auto ML*



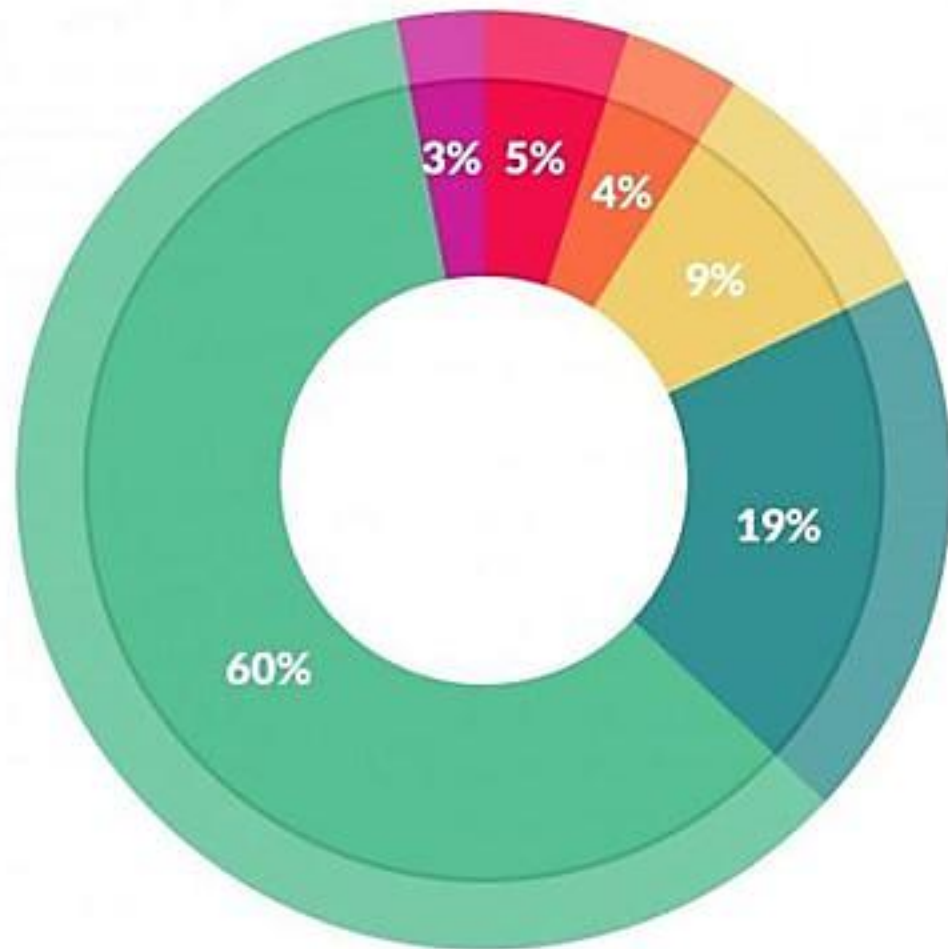
# Poderia uma IA usar IA?

## *AutoML*

- Inteligência Artificial apoiando na criação de modelos de *Machine Learning (ML)*
- Redução o tempo na criação de modelos de ML
- Diminui a barra de conhecimento necessário em ML
- Será o fim da carreira dos cientistas de dados?



# Data Science Report



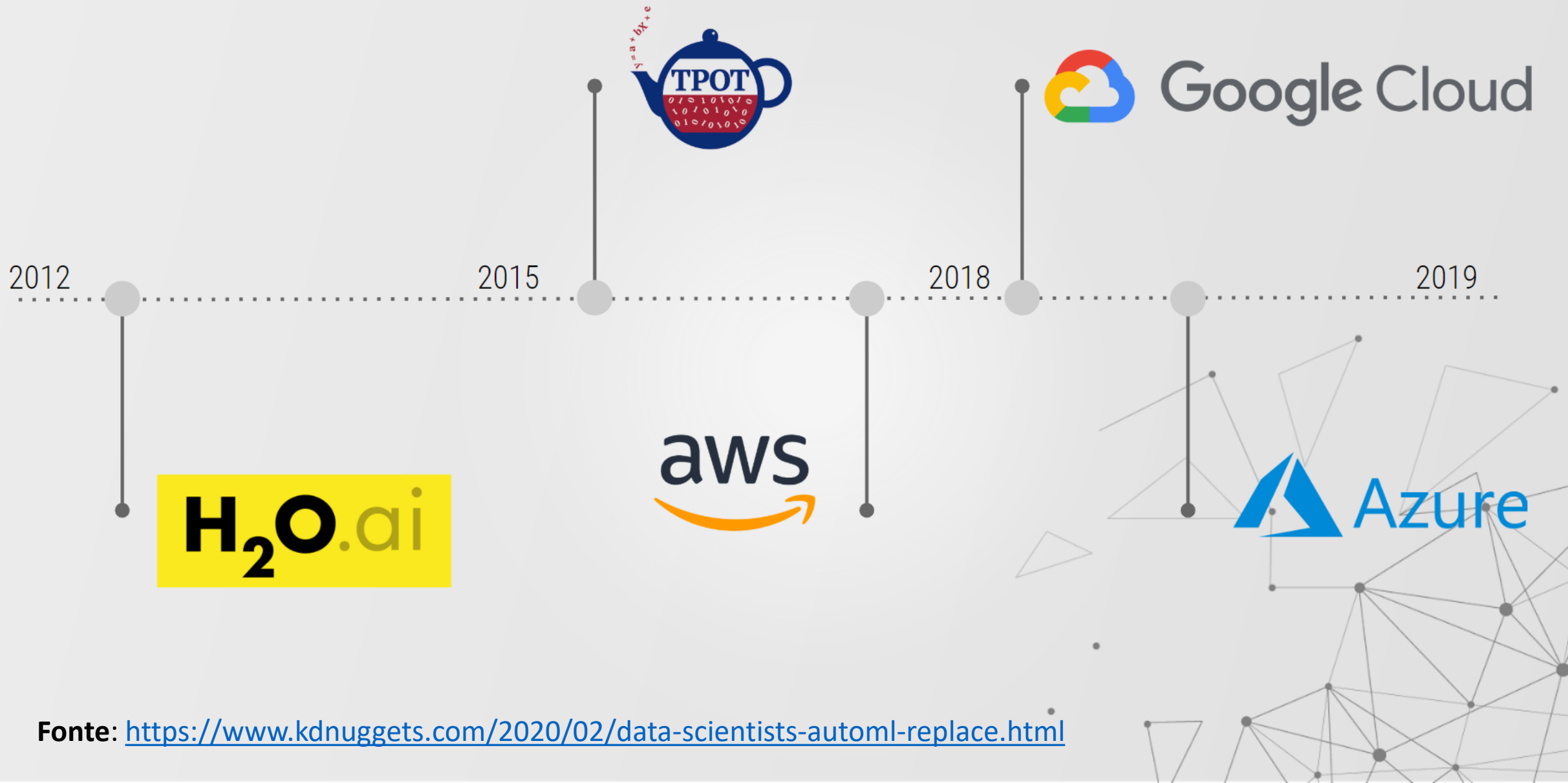
## What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

> **80%** do tempo em tarefas relacionadas a pré-processamento



# Tecnologias e Players



Fonte: <https://www.kdnuggets.com/2020/02/data-scientists-automl-replace.html>

# AutoML

## Principais etapas

- *Features selection/Engineering*
- Otimização de hiper parâmetros
- Comparação de modelo

## Acesso fácil e simplificado para *non-experts*

- Google Cloud AutoML (lançado em 2018)
- Microsoft Azure AutoML (lançado em 2018)
- Amazon SageMaker (lançado em 2017)

# AutoML

## *H2O.ai*

- Lançado em 2012 oferece bibliotecas (*open-source*). Em 2017 lança o serviço comercial, chamado *Driverless AI*. Adotado em várias indústrias, como serviços financeiros e varejo.

## TPOT (Tree-based Pipeline Optimization Tool)

- Biblioteca Python *free* implementa algoritmo genéticos para “evoluir e identificar” melhores modelos. Desenvolvido pela Universidade da Pensilvânia (EUA). Alcança 97% de precisão no famoso *dataset* IRIS.
- Pode levar horas/dias processando...

# MLOps

## Evolução

- Revolução industrial: surgimento de máquinas físicas exigiu que as organizações sistematizassem suas operações, formando fábricas, linhas de montagem e tudo o que sabemos sobre manufatura automatizada.
- Sistemas *Agile* ajudaram organizações a operacionalizar ciclo de vida de produtos, favorecendo a inovação contínua.
- *Devops* otimizou ainda mais o ciclo de vida da produção e favoreceu o *Big Data*.
- Chegou a horas de organizar o *Machine Learning*



# MLOps

## Definição

- Prática de colaboração e comunicação entre cientistas de dados e profissionais de operação
- Visa auxiliar a gestão do ciclo de vida de processos de *Machine Learning*
- Aumento da automação e melhoria da qualidade de produção de ML

## Aplicação

- Foca nos requisitos de negócios e regulação
- Conjunto de boas práticas – elimina o desperdício, melhora qualidade
- Deve ser independente de "plataforma"
- Facilita a reprodutibilidade de experimentos
- Gestão de artefatos (*scripts*, bases de dados, gráficos, etc)

# *MLOps – Um exemplo*

## *Data Version Control · DVC*

- Open-source Version Control System for Machine Learning Projects <https://dvc.org/>

## O que oferece

- Controle de versionamento de projeto (modelos e dados)
- Gestão de experimentos (reprodutibilidade e consistência)
- Implantação e colaboração



# **PUC Minas Virtual**