

# MIOps

---

Advanced Institute for Artificial Intelligence – AI2

<https://advancedinstitute.ai>

## Agenda

- ☐ MLOps
- ☐ Processo
- ☐ PyCaret
- ☐ MLflow

## MLOps <sup>1</sup>

- ❑ Convergência entre engenharia de software, DevOps e Machine Learning
- ❑ Alguns desafios de projetos de Machine Learning:
  - Colocar em operação
  - Avaliar o desempenho
  - Evoluir continuamente
  - Inspeccionar o modelo

---

<sup>1</sup>Manifesto MLOps: <https://mlops.community/manifesto/>

Dificuldades enfrentadas por equipes multidisciplinares ao desenvolverem projetos de machine learning

- ☐ Não é possível implantar - modelos bloqueados
- ☐ Colaboração ineficiente
- ☐ Rastreamento manual
- ☐ Sem reprodutibilidade ou proveniência
- ☐ Modelos não monitorados

## Proveniência de dados

- ❑ Gerencia todo o processo de transformação de dados desde a sua concepção
- ❑ Permite identificar quem criou, quem alterou, qual motivação para alteração, entre outros
  - Exemplo: a base de exemplo de imagens aumentou. Pode ser importante registrar o id do experimento que resultou em novos dados. E o id remeter a uma informação completa a cerca do experimento
- ❑ Modelos de machine learning são fortemente dependente dos dados. Se os dados se alteram sem o cientista de dados ter conhecimento, o modelo vai apresentar resultados diferentes
  - A proveniência tem um papel fundamental no desenvolvimento de modelos de machine learning

## Requisitos para atingir maturidade no processo de desenvolver modelos

- ❑ Reprodutibilidade
  - Mesmo na área acadêmica, poucas experiências são facilmente reproduzíveis
- ❑ Colaboração
  - Colaboração assíncrona é fundamental para que um modelo evolua levando em conta o conhecimento de um grupo
- ❑ Accountability
  - Comparar os resultados de um modelo em produção com o resultado na prototipação
- ❑ Desenvolvimento contínuo
  - Monta o modelo automaticamente e avaliar estatisticamente

## Inspecionando construção de modelos

- ❑ As ferramentas de desenvolvimento de software se concentram em commits de código. Mas ML tem outros aspectos que precisam ser avaliados: dados, modelos e métricas.
- ❑ Necessário rastrear execuções, dados agrupados, versões de código e parâmetros que foram usados, para criar um conjunto de dados intermediário ou um modelo.
- ❑ Isso fornece contexto completo para reprodutibilidade e proveniência para conectar a engenharia de dados com o treinamento do modelo para manter accountability.

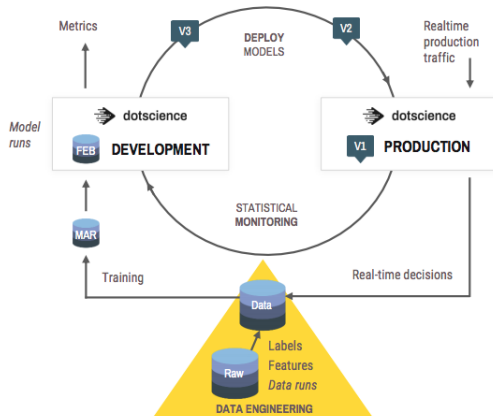
O Desenvolvimento de um modelo é uma atividade desafiadora e complexa, mesmo considerando que em geral parte de um modelo similar existente

- Dividir o desenvolvimento em termos de atividades menores não é trivial
- A concepção do modelo passa por etapas de colaboração em um time multi-disciplinar
- Partes desse processo são mais sistemáticas e podem ser automatizadas
  - Colocar um dado modelo em produção
  - Comparar desempenho de diferentes modelos



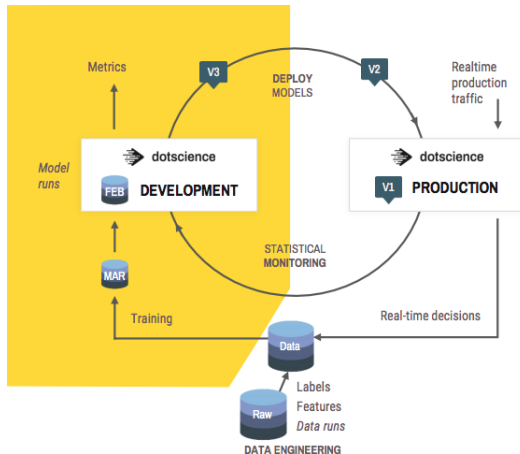
# Desenvolvimento Contínuo

Na etapa de preparação de dados é necessário registrar as transformações. De forma que possam ser reproduzidas do dados original, até o dados preparado para o modelo. <sup>2</sup>



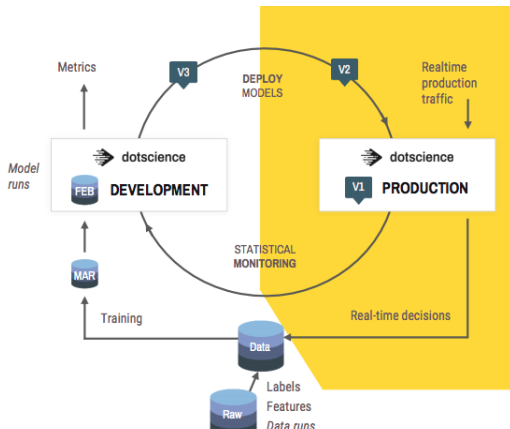
# Desenvolvimento Contínuo

Compartilhamento do código com documentação que permita reprodução. Assim, outras pessoas podem integrar a equipe e continuar o trabalho do ponto onde parou.<sup>3</sup>



# Desenvolvimento Contínuo

Ao colocar em produção é necessário disponibilidade e montado de forma automática e manter com desempenho adequado. Para isso é necessário emitir alertas a respeito do monitoramento estatístico durante o uso do modelo.<sup>4</sup>



## CD4ML (Continuous Delivery for Machine Learning) e MLOps


- Nesse cenário a integração e entrega contínua nesse processo é fundamental, pois a comunicação entre Cientistas de Dados e a equipe de operações ou produção é fundamentalmente colaborativa
- Tal colaboração precisa ser automatizada para colocar sistemas de aprendizagem de máquina em produção mais rápido e minimizar os riscos
- CD4ML é mais focada na entrega contínua enquanto que MLOps é mais focado na operacionalização dos modelos



## Algumas ferramentas para CD4ML e MLOPs



- ☐ Uber Michelangelo
- ☐ FBLearner
- ☐ Tensorflow TFX
- ☐ Apache MLFlow
- ☐ Pycaret
- ☐ AutoML

O MLFlow é um exemplo de ferramenta que automatiza e facilita o controle do desenvolvimento de aplicações de aprendizagem de máquina

- O MLflow apresenta as seguintes funcionalidades:
  - Permite armazenar métricas de qualidade de diversos modelos relacionados a um mesmo experimento
  - Permite gerar o executável do modelo a partir de diversos frameworks


 **Experiments** **Models** [GitHub](#) [Docs](#)

**Experiments**  



**Training**  


## Training




Experiment ID : 0 Artifact Location : file:///data/output/mlruns/0



▼ **Notes** 

None

Search Runs:   State: **Active**  **Search** **Clear**

Showing 18 matching runs **Compare** **Delete** **Download CSV** 

   **Columns**

						Parameters			Metrics >		
<input type="checkbox"/>	Start Time	Run Name	User	Source	Version	alpha	k	l1_ratio	Test scor	Test_scor	acc
<input type="checkbox"/>	✓ 2020-07-08 11:41	-	silvio	 fashior	e232e3	-	-	-	-	-	0.866
<input type="checkbox"/>	✓ 2020-07-08 11:41	-	silvio	 fashior	e232e3	-	-	-	-	-	0.868

## Funcionalidades do MLFLOW

- ❑ Tracking: monitorar experimentos de modelos de aprendizagem de máquina
- ❑ Project: reprodução de experimentos
- ❑ Models: automaticamente gerar um modelo que pode ser montado em diversas plataformas



## Tracking

- ☐ Parâmetros do modelo
- ☐ Métricas de avaliação de modelos
- ☐ Artefatos diversos relacionados ao modelo
- ☐ Código fonte

## Reprodução

- ☐ Local ou Remota
- ☐ Definição de ambiente (conda.yaml)
- ☐ Definição dos parâmetros da execução do modelo (MLProject)

## Modelo

- Controla a montagem do modelo em diversas plataformas
  - Azure, Aws, datbricks, etc

## PyCaret <sup>5</sup>

- Ferramenta para criar modelo de modo automatizado
  - Avalia dados de entrada
  - Compara algoritmos
  - Otimiza parâmetros
  - Avalia o resultado
  - Módulo de interpretação para modelos de árvore

---

<sup>5</sup> [pycaret.org](https://pycaret.org)