

# MLFLOW Aplicado à Soluções de Machine Learning

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Inteligência Artificial e Aprendizado de Máquina  
por Nicolas Marcos de Moraes Oliveira

# Introdução

- Objetivo: Apresentar uso da plataforma de MLFLOW em soluções corporativas de Machine Learning.
  - Implementação através de ambiente Databricks;
  - Estudo de caso com base fictícia de RH, com problema de classificação com objetivo de analisar a probabilidade de um funcionário deixar a empresa.

# Motivação

- Conforme afirmado por Géron (2019, p. 24), ao trabalhar com Machine Learning e partir do princípio que a principal tarefa consiste em selecionar o algoritmo de aprendizagem e o treinar baseado em dados, dois possíveis pontos de erro são algoritmos ruins e dados ruins.
- Corroborado por Vanschoren e Blockeel (2009, p.1), experimentação é alma de Machine Learning.



# Sobre o MLFLOW

## MLflow Components

### mlflow Tracking

Record and query  
experiments: code,  
configs, results,  
...etc

### mlflow Projects

Packaging format  
for reproducible  
runs  
on any platform

### mlflow Models

General model format  
that supports diverse  
deployment tools

# Sobre os Dados

COLUNA	DESCRIÇÃO	TIPO
Age	Idade do funcionário	Numérico discreto
Attrition	Se o funcionário deixou ou não a companhia: YES / NO	Catégorico nominal
BusinessTravel	Frequência com a qual o funcionário viaja	Catégorico nominal
DailyRate	Taxa diária do funcionário	Numérico discreto
Department	Departamento do funcionário	Catégorico nominal
DistanceFromHome	Distância entre a moradia do funcionário e seu trabalho	Numérico discreto
Education	Nível educacional do funcionário, sendo: 1 'Below College'; 2 'College'; 3 'Bachelor'; 4 'Master'; 5 'Doctor'.	Numérico discreto (representando catégorico ordinal)
EducationField	Área de estudo do funcionário.	Catégorico nominal
EmployeeCount	Valor dummy para contagem de registros	Numérico discreto
EmployeeNumber	Número identificador do funcionário	Numérico discreto
EnvironmentSatisfaction	Nível de satisfação do funcionário com ambiente de trabalho, sendo: 1 'Low'; 2 'Medium'; 3 'High'; 4 'Very High'.	Numérico discreto (representando catégorico ordinal)
Gender	Gênero do funcionário	Catégorico nominal
HourlyRate	Taxa hora do funcionário	Numérico discreto
JobInvolvement	Nível de envolvimento do funcionário com o trabalho, sendo: 1 'Low', 2 'Medium', 3 'High', 4 'Very High'.	Numérico discreto (representando catégorico ordinal)
JobLevel	Nível do funcionário	Numérico discreto
JobRole	Cargo do funcionário	Catégorico nominal
JobSatisfaction	Nível de satisfação do funcionário com o trabalho, sendo: 1 'Low', 2 'Medium',3 'High', 4 'Very High'.	Numérico discreto (representando catégorico ordinal)
MaritalStatus	Status civil do funcionário	Catégorico nominal
MonthlyIncome	Rendimento mensal do funcionário	Numérico discreto
MonthlyRate	Taxa mensal do funcionário	Numérico discreto
NumCompaniesWorked	Quantidade de companhias onde o funcionário já trabalhou	Numérico discreto
Over18	Se “Y” referente à “yes”, indica que o funcionário é maior de 18 anos	Catégorico nominal

# Sobre os Dados

COLUNA	DESCRIÇÃO	TIPO
OverTime	Indica se o funcionário faz horas extras como “Yes” ou “No”	Categórico nominal
PercentSalaryHike	Evolução salarial percentual	Numérico discreto
PerformanceRating	Nível de performance do funcionário, sendo: 1 'Low', 2 'Good', 3 'Excellent', 4 'Outstanding'	Numérico discreto (representando categórico ordinal)
RelationshipSatisfaction	Nível de satisfação do funcionário em relacionamento, sendo: 1 'Low', 2 'Medium', 3 'High', 4 'Very High'.	Numérico discreto (representando categórico ordinal)
StandardHours	Quantidade de horas padrões do funcionário	Numérico discreto
StockOptionLevel	Nível de opção de participação em ações	Numérico discreto
TotalWorkingYears	Total de anos trabalhados	Numérico discreto
TrainingTimesLastYear	Vezes em que o funcionário foi treinado no último ano	Numérico discreto
WorkLifeBalance	Nível de equilíbrio entre vida pessoal e trabalho, sendo: 1 'Bad', 2 'Good', 3 'Better', 4 'Best'.	Numérico discreto (representando categórico ordinal)
YearsAtCompany	Anos do funcionário na companhia	Numérico discreto
YearsInCurrentRole	Anos do funcionário no cargo atual	Numérico discreto
YearsSinceLastPromotion	Anos do funcionário desde sua última promoção	Numérico discreto
YearsWithCurrManager	Anos do funcionário com gestor atual	Numérico discreto

\*Observação: O dataset não dispunha de metadados a respeito de todas as colunas. Portanto, as descrições citadas são a respeito do entendimento do autor sobre as colunas, corroborados por pesquisas em canais e fóruns sobre estudos realizados nos mesmos dados.

# Continuação da Apresentação no Código! =D

- Projeto Disponível em:  
[https://github.com/nicolasmarcos/TCC\\_PucMinas](https://github.com/nicolasmarcos/TCC_PucMinas)