

Big Data Analytics com R e Microsoft Azure Machine Learning



Data Science Academy



O que é o Microsoft Azure Machine Learning?



Data Science Academy

www.datascienceacademy.com.br



O que é o Microsoft Azure Machine Learning?

Microsoft Azure Machine Learning é um framework de Machine Learning que pode ser usado para construir modelos de análise preditiva, através de conjuntos de dados de treinamento e uma variedade de fontes de dados



Data Science Academy



O que é o Microsoft Azure Machine Learning?

Machine Learning em nuvem (cloud)



Data Science Academy




O Microsoft Azure Machine Learning oferece uma forma rápida de trabalhar com análise preditiva



Data Science Academy

www.datascienceacademy.com.br

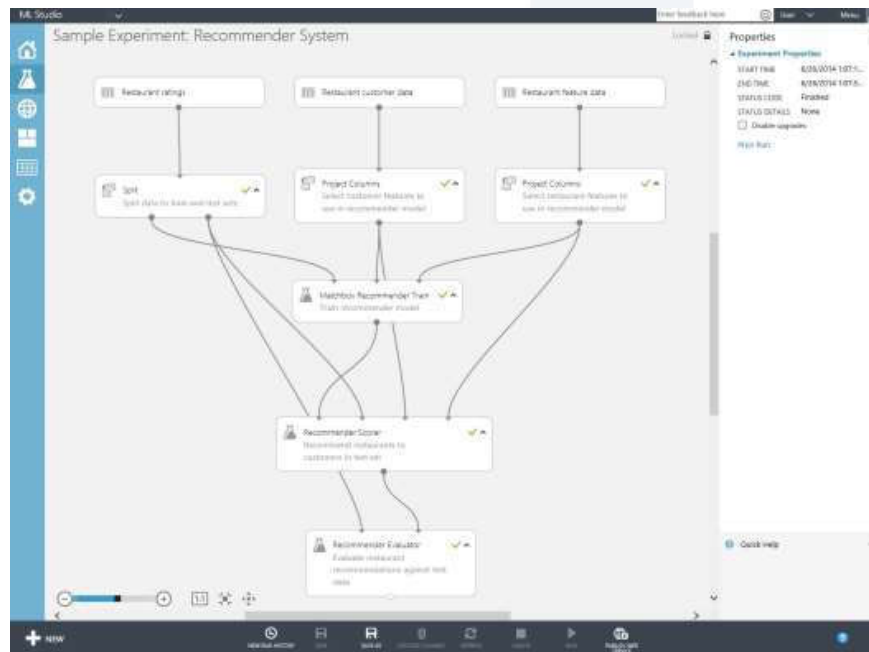


Análise Preditiva é a tecnologia por trás do Microsoft Azure Machine Learning e pode simplesmente ser definida como uma forma científica de usar dados passados para prever o futuro



Data Science Academy

Azure Machine Learning Studio



- R
- Python
- SQL



Data Science Academy

www.datascienceacademy.com.br



O Azure ML possui 3 grandes concorrentes

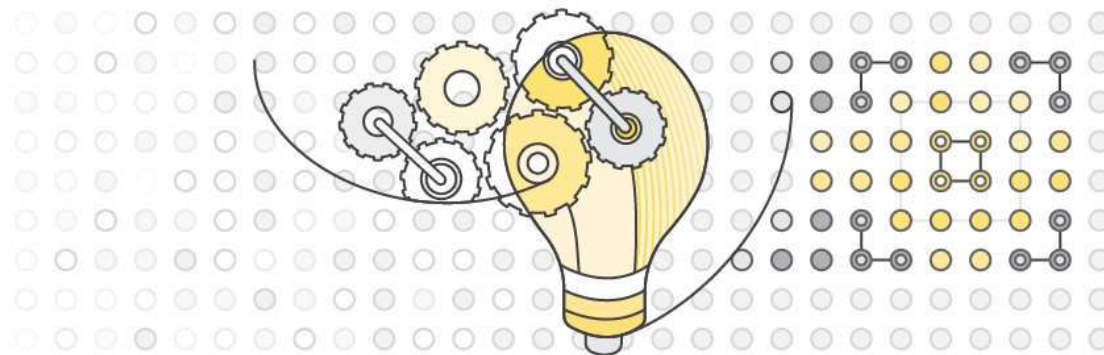


Data Science Academy

www.datascienceacademy.com.br



Amazon Machine Learning

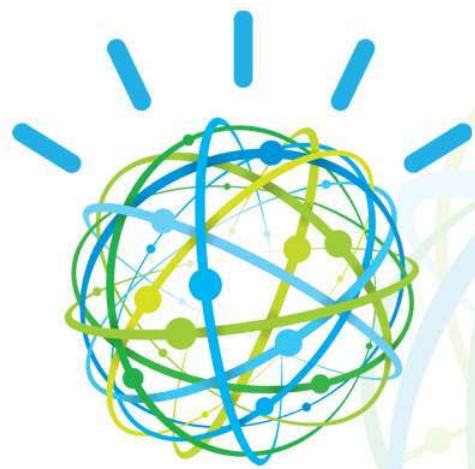


Data Science Academy

www.datascienceacademy.com.br



IBM Watson Analytics



**IBM
Watson
Analytics**

AaaS



Data Science Academy

www.datascienceacademy.com.br



Apache Spark



Data Science Academy

www.datascienceacademy.com.br

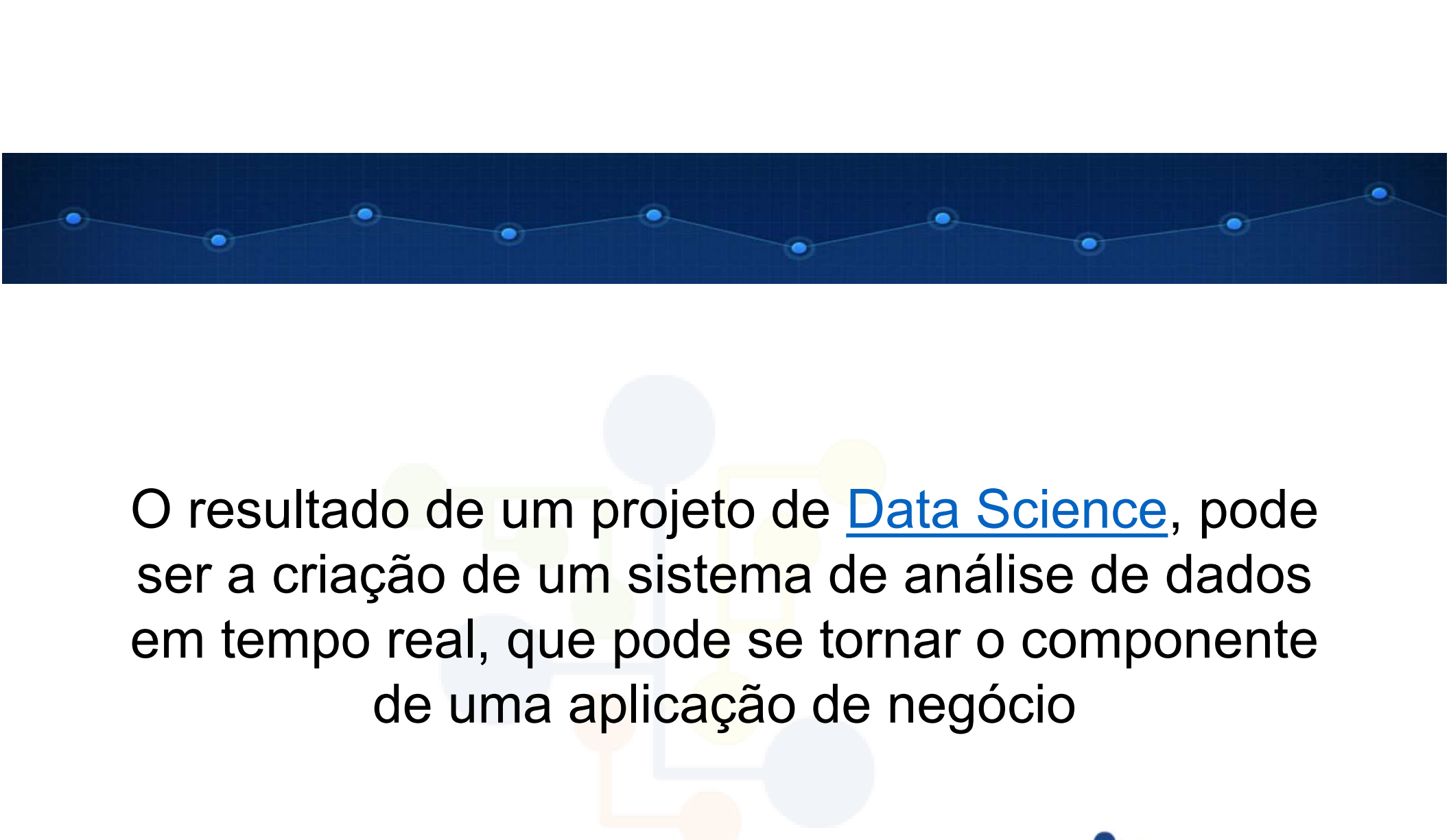


Apache Spark

Spark suporta diferentes tarefas de [Data Science](#)



Data Science Academy



O resultado de um projeto de Data Science, pode ser a criação de um sistema de análise de dados em tempo real, que pode se tornar o componente de uma aplicação de negócio



Data Science Academy

Azure Machine Learning Workflow



Data Science Academy



Azure Machine Learning Tool Kit

- **Classificação** - algoritmos usados para classificar dados em diferentes categorias e então fazer previsões sobre dados com variáveis qualitativas.
- **Regressão** - algoritmos usados para prever uma ou mais variáveis numéricas.
- **Clustering** - algoritmos usados para identificar grupos ou padrões em conjuntos de dados e então fazer previsões de classificação em grupos, para uma determinada variável.



Data Science Academy

Criar um Novo Modelo



Testar o Novo Modelo



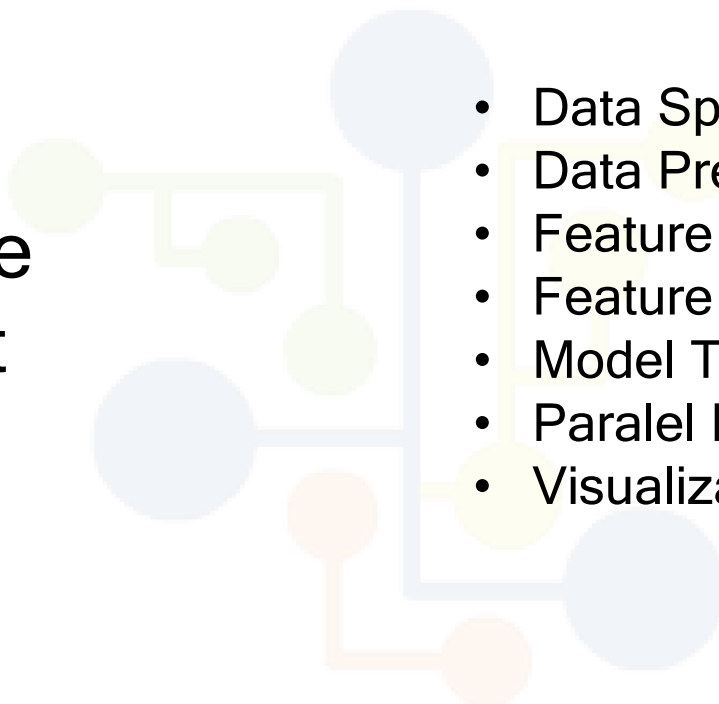
Deploy do Novo Modelo



Data Science Academy



Pacote Caret

- 
- Data Splitting
 - Data Pre-processing
 - Feature Selection
 - Feature Importance
 - Model Tuning
 - Parallel Processing
 - Visualization



Data Science Academy