4·Linux Open Software Specialists™

CERTIFICADO

Certificate

Rafael Batista

Concluiu seus estudos no curso **522 - Python and Spark for Big Data and Machine Learning - AC** ministrado pela empresa 4Linux e cumprindo a carga horária de 40 horas.

18 de outubro de 2026



RODOLFO GOBBI

DIRETOR GERAL

Para validar a autenticidade deste certificado acesse aia.4linux.com.br/admin/tool/certificate/ e digite o código: 0189642014RB

Ementa de Curso

Rafael Batista

Concluiu seus estudos no curso **522 - Python and Spark for Big Data and Machine Learning - AC** ministrado pela empresa 4Linux e cumprindo a carga horária de 40 horas.
18 de outubro de 2026

A autenticidade deste documento pode ser verificado em aia.4linux.com.br/admin/tool/certificate/ digitando o código 0189642014RB

Conteúdo Programático

Introdução a Big Data

- Python para Big Data
- Jupyter Notebook
- Google Colab

Python para Big Data

- O modelo de programação MapReduce
- Introdução à Programação funcional com Python: Funções Lambda em Python
- Trabalhando com grafos em Python
- Numpy
- Pandas

Análise exploratória de Dados

- Tipos de Variáveis
- Normalização de Dados com Python
- Visualização de Dados com Python (Matplotlib, Searborn, Bokeh)
- Medidas de similaridade

Matemática e Dados

- Introdução à Estatística Descritiva
- Média, Mediana, Desvio Padrão, Variância
- Teorema de Bayes e Inferência Bayesiana
- Teste Estatístico de Hipótese
- Exemplo: Executando um Teste A/B

Spark e Pyspark

- Operações básicas com DataFrames
- Operações avançadas com DataFrames
- Operações com Dados Faltantes e Datas

Introdução ao Machine Learning

- Introdução a Machine Learning
- Overview dos Algoritmos de Machine Learning
- K-nearest neighbors – kNN
- Construção de Pipeline de Machine Learning com modelo CRISP DM
- Introdução a Sistemas de Recomendação e Filtros Colaborativos
- **Machine Learning**
- Regressão Linear
- Regressão Logística x kNN – Avaliação de Modelos
- Fundamentos de Redes Neurais Artificiais
- Regressão Logística x kNN – Avaliação de Modelos
- Introdução a Processamento Natural de Linguagem (Modelos n-grams)
- Word2Vec com Gensim
- Clusterização usando Kmeans
- Classificação usando Naive Bayes
- Classificação usando Árvores de Decisão e Random Forest
- Classificação usando Support Vector Machine
- **Introdução ao Machine Learning Spark MLlib**
- Introdução ao Spark Mllib

- Tokenização de documentos de textos para tarefas de classificação
- Regressão Logística com Spark Mllib
- KNN com Spark Mllib
- Spark Streaming

- **Deep Learning**

- Introdução a Deep Learning
- Redes Neurais Convolucionais
- Redes Neurais Recorrentes
- Introdução ao TensorFlow
- Conhecendo a base de dados MNIST
- Aplicação Prática de Redes Neurais MLP para Reconhecimento de Imagem
- Aplicação Prática de Redes Neurais Convolucionais para Reconhecimento de Imagem
- Aplicação Prática de Redes Neurais Recorrentes para Análise de Sentimentos

- **Processamento de Alto Desempenho com Python para Big Data**

- A revolução das GPUs em Machine Learning
- Arquitetura Cuda e Modelo de Programação em GPU
- Introdução ao Numba
- Comparação de desempenho entre Numpy, Numpy assíncrono e Numpy com Numba em Operações estatísticas
- Conhecendo os principais frameworks Python para Machine Learning que usam GPUs