



Python:

TOP 10 BIBLIOTECAS PARA DATA SCIENCE

BY ALEXANDRA RAIBOLT

1.

NumPy

Alicerce para **COMPUTAÇÃO CIENTÍFICA** apresenta objetos de grandes **MATRIZES** e matrizes **MULTIDIMENSIONAIS**, além de funções **MATEMÁTICAS** e **ESTATÍSTICAS** de alto nível.

Outras bibliotecas de nível superior listadas aqui, utilizam **NUMPY** como suas entradas e saídas.

Baseado no NumPy,
ESTENDENDO seus
recursos.

Acrescenta algoritmos e
comandos para **MANIPULAR**
e **VISUALIZAR** dados.

Inclui **DIVERSOS MÓDULOS**
tais como álgebra linear,
teoria da probabilidade,
otimização numérica,
integração, etc.,
POTENCIALIZANDO tarefas
de ciência de dados.



SciPy

3.

Matplotlib

Biblioteca **PADRÃO** para visualização em Python, portanto, **PERMITE** a criação de **GRÁFICOS** e **PLOTAGENS 2D** em nível profissional.

OUTRAS bibliotecas de **VISUALIZAÇÃO** são criadas para operar em **CONJUNTO** com o Matplotlib, trazendo **RECURSOS** interativos, **EXPORTAÇÃO** de gráficos, etc.

Baseada no Matplotlib
contém recursos de ALTO
NÍVEL para
PROCESSAMENTO de
gráficos.

Possui TIPOS de
visualizações mais
COMPLEXAS como
diagramas conjuntos, séries
temporais, etc.



Seaborn

5.

Scrapy

Usada para a criação de **BOTS SPIDERS** com a finalidade de **RASTREAR** páginas **WEB** e coletar **DADOS** estruturados.

Tais dados podem ser URLs, preços, contatos, etc., além de ter a **CAPACIDADE** também de coletar dados de **APIs**.

Projetada para
ESPECIFICAMENTE
problemas de Processamento
de Linguagem Natural (PLN).

Permite **MARCAÇÃO** de
texto, **ÁRVORES** de análise,
IDENTIFICAÇÃO de entidades
nomeadas, etc.

Tais **FUNCIONALIDADES**
permitem a realização, por
exemplo, de **ANALISE** de
SENTIMENTOS.



NLTK



7.

SciKit-Learn

Baseado no NumPy e SciPy,
ADICIONA uma quantidade
enorme de ALGORITMOS
para tarefas de
APRENDIZADO DE
MÁQUINA e MINERAÇÃO
DE DADOS.

Inclui por exemplo TÉCNICAS
de clusterização, classificação
e regressão.

Desenvolvido pelo **GOOGLE**,
usado **ESPECIFICAMENTE**
para **APRENDIZADO DE**
MÁQUINA.

Com ele, é possível
implementar, modelar e
treinar **DIVERSOS** tipos de
arquiteturas de **REDES**
NEURAS ARTIFICIAIS
utilizando uma quantidade
enorme de conjunto de
DADOS.



TensorFlow



Keras

Pode ser rodada sobre **TENSORFLOW** ou **THEANO** (Top 10), além de poder utilizar **OUTRAS** bibliotecas não listadas nesta lista (CNTK, MxNet) como **BACKENDS**.

Trabalha especificamente com **REDES NEURAIAS**.

Utiliza a **SINTAXE** NumPy para **OTIMIZAR** expressões matemáticas.

Aproveita **GPUs** para realizar **CÁLCULOS** em até **100** vezes mais **RÁPIDO** do que **CPUs**.

Extremamente **VANTAJOSO** na utilização de tarefas de **DEEP LEARNING** e tarefas computacionais **COMPLEXAS**, devido a sua **VELOCIDADE** de **PROCESSAMENTO**.



Theano

ALEXANDRA RAIBOLT

[Site](#) | [GitHub](#) | [LinkedIn](#) | [Lattes](#) | [E-mail](#)

<Sistema FIRJAN>

Estagiária de BI e Administração de Dados.

</>

<FAETERJ>

Cursando Tecnologia da Informação & Comunicação.

</>

<Estácio de Sá>

Cursando Sistemas da Informação.

</>