

Các thư viện Python phổ biến trong Khoa học Dữ liệu 2017 (Phần 1)

Lành Nguyễn

Ngày 8 tháng 9 năm 2017

Tóm tắt nội dung

Python là thư viện được sử dụng khá phổ biến trong lĩnh vực khoa học dữ liệu. Bài viết này liệt kê danh sách các thư viện hữu ích được dùng trong phân tích dữ liệu. Qua đó, chúng ta có cái nhìn tổng thể hơn về các thư viện này. Các thư viện được chia thành 4 nhóm: Core libraries, Visualization, Artificial Intelligence (Machine Learning, Deep Learning, Natural Language Processing) và Data Mining - Statisticis.

1 Core Libraries

Bao gồm các thư viện: Numpy, Scipy, và Pandas.

Fig. 1.

Hình 1: Every plot must have axes labeled.

1.1 Numpy

1.2 Scipy

1.3 Pandas

Trong phần tiếp theo - Phần 2, chúng tôi sẽ giới thiệu các thư viện trực quan hoá dữ liệu (Visualization) bằng biểu đồ, đồ thị bao gồm *Matplotlib*, *Seaborn*, *Koreh*, *Plotly*.

[Top 15 Python Libraries for Data Science in 2017](#)

Tài liệu

- [1] A. C. Melissinos and J. Napolitano, *Experiments in Modern Physics*, (Academic Press, New York, 2003).

- [2] N. Cyr, M. Têtu, and M. Breton, IEEE Trans. Instrum. Meas. **42**, 640 (1993).
- [3] *Expected value*, available at http://en.wikipedia.org/wiki/Expected_value.