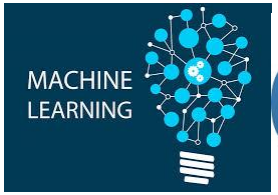
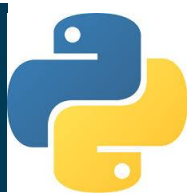
Trường ĐH Khoa Học Tự Nhiên Tp. Hồ Chí Minh
TRUNG TÂM TIN HỌC

BIG DATA IN MACHINE LEARNING

Giảng viên: Khuất Thùy Phương

2020





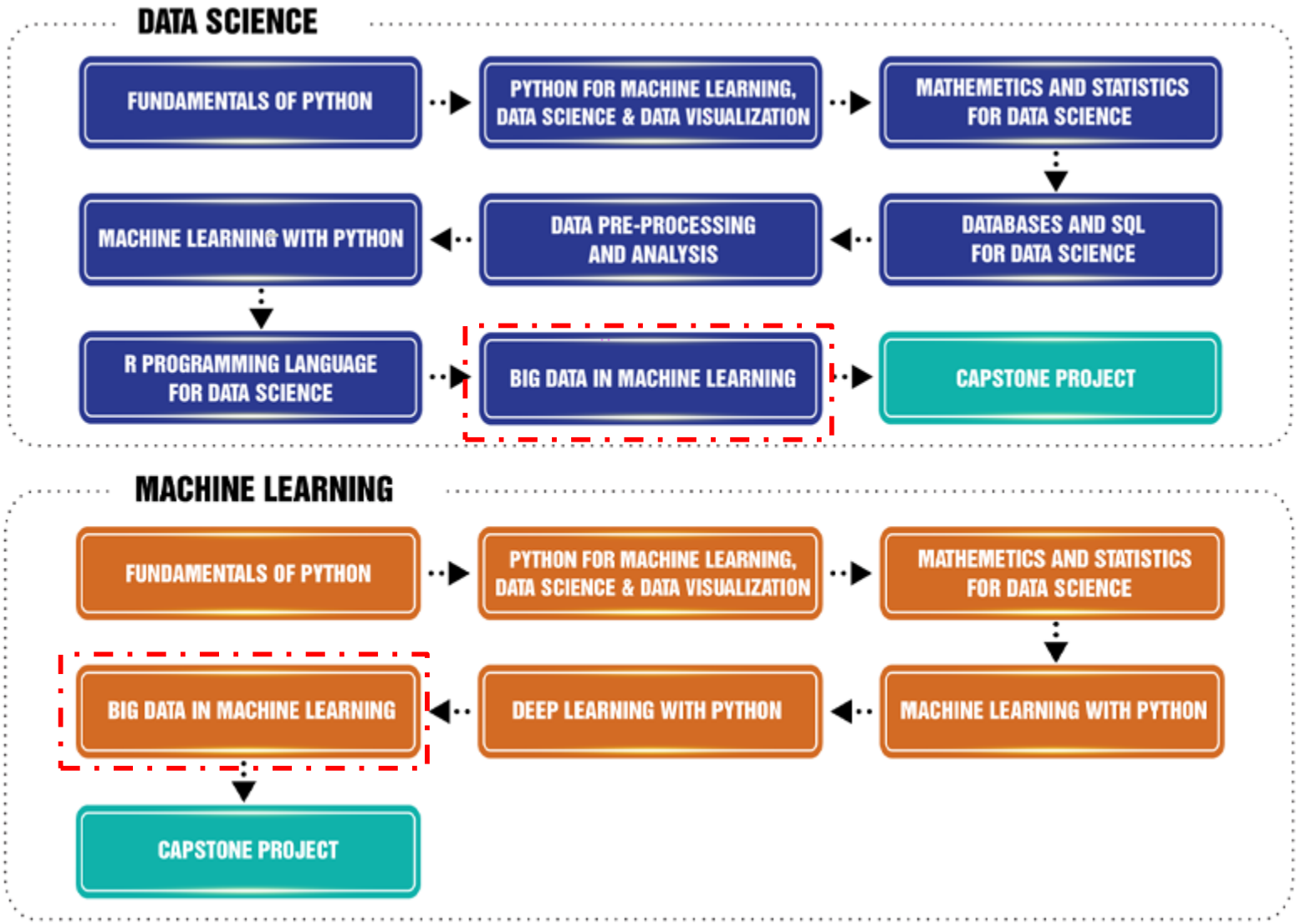
BIG DATA IN MACHINE LEARNING

- **Hình thức học:** học online 2 buổi/tuần
  - **Thời gian:** 21/03 – 05/04 (từ **06/04:** sẽ có thông báo sau)
  - **Lịch học : Thứ Bảy & Chủ Nhật** (13h30 – 17h30)
- **Đánh giá môn học:**
  - **50%:** chuyên cần tham gia các buổi online + hoàn thành bài tập hàng tuần (gửi bài nộp qua email) – Subject email: LDS9_BT_Tuan_X_HoTen
  - **50%: L**àm đồ án
- **Công cụ hỗ trợ: Zoom**
- **Phụ trách:**
  - **Khuất Thùy Phương**
  - Email: tubirona@gmail.com

Lớp học online

Học trực tiếp với GV theo lịch

R PROGRAMMING LANGUAGE FOR DATA SCIENCE



Nội dung

- Overview of Big Data
- Overview of PySpark
- PySpark RDDs
- PySpark SQL and DataFrame
- Data Preprocessing & Analysis
- Overview of PySpark Mllib

Nội dung

- **Supervised Learning (Classification & Regression)**
  - Linear Regression (pyspark.ml.regression)
  - Logistic Regression (pyspark.ml.classification)
  - Decision Tree (pyspark.ml.classification)
  - Random forest (pyspark.ml.classification)
  - Gradient-Boosted Trees (pyspark.ml.classification)
  - Pipeline

Nội dung

- **Unsupervised Learning (Clustering & Recommender System)**
  - Clustering with K-means (pyspark.ml.clustering)
  - Recommender System (pyspark.ml.recommendation)

Nội dung

- **Unsupervised Learning (Clustering & Recommender System)**
  - Clustering with K-means (pyspark.ml.clustering)
  - Recommender System (pyspark.ml.recommendation)

Nội dung

- PySpark Streaming
- Natural Language Processing (NLP)
- Apache Spark standalone cluster