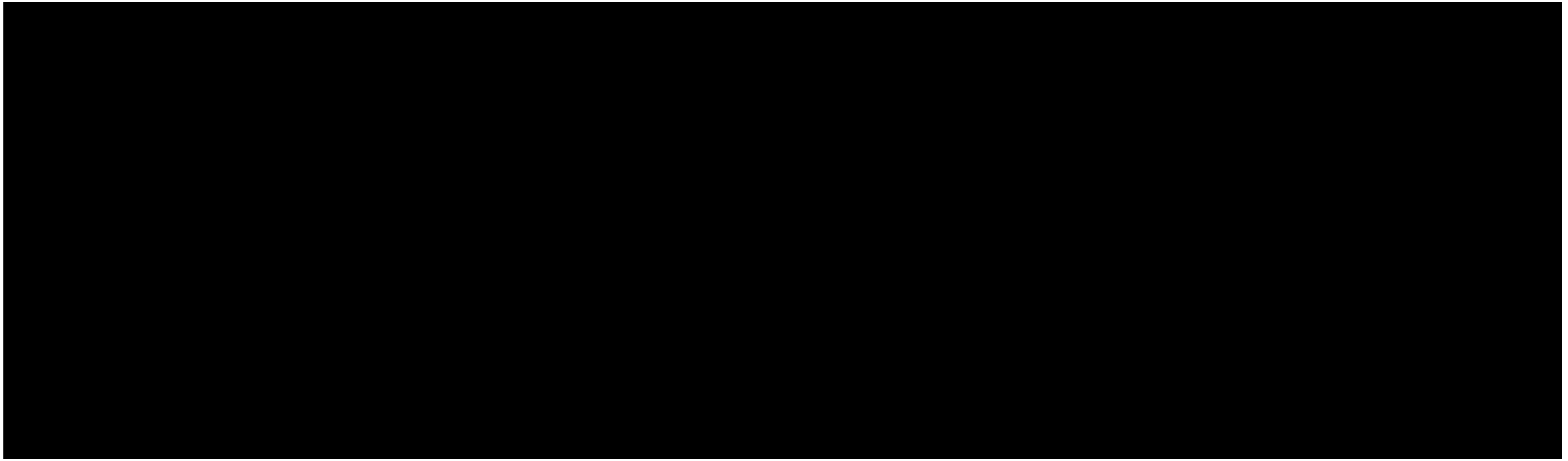
INTRODUCTION

DATA & ARTIFICIAL INTELLIGENCE/MACHINE LEARNING @ DIGITAL ACADEMY



THE DATA SCIENCE TSUNAMI

With the advent of Deep Learning and Big Data computing, **data science teams** are fast becoming a must-have for decision-makers.

Harvard Business Review's announcement of data scientists as the rock stars of the 21st Century has led to heightened interest in the field.

Will the hype have subsided 5 years from now?

Why are you here?

IDEAL SCENARIO

Masters in Data Science

- 8 courses over 3-4 terms, with examinations and assignments
- **200** hours in the classroom + **1000** hours of practical work
- **core courses** (5)
  - basics (math/stats, programming, databases, communication, data processing)
  - introduction to data science (survey course)
  - big data mining
  - text mining and natural language processing
  - putting it all together (project-based course)
- **options** (6 half-semester courses)

focus on classification
focus on clustering
deep learning I and II
data streams
multi-media mining
advanced visualization

recommender systems
social network analysis
outlier detection
event mining
frequent pattern mining
etc.

Session			AI/ML Stream	Data Stream
1	31-Jan-19	1-4 PM	Data Science Universals	
2*	7-Feb-19	9-12 AM	*NA	Basics of Programming - Fundamental Concepts
2	7-Feb-19	1-4 PM	Basics of Programming with R and Python	
3	14-Feb-19	1-4 PM	Statistical and Mathematical Foundations	Data + Information Architecture
4	21-Feb-19	1-4 PM	Data Collection and Data Processing	
5	28-Feb-19	1-4 PM	Data Exploration and Data Visualization	
6	7-Mar-19	1-4 PM	Statistical Learning; Association Rules Mining	Basic Data Analytics Techniques
7	14-Mar-19	1-4 PM	Lab: Association Rules Mining	Lab: Basic Data Analytics Techniques
8	21-Mar-19	1-4 PM	Classification + Value Estimation	Measures and Metrics
9	28-Mar-19	1-4 PM	Lab: Classification + Value Estimation	Lab: Measures and Metrics
10	4-Apr-19	1-4 PM	Clustering	Advanced Data Visualization and Reporting
11	11-Apr-19	1-4 PM	Lab: Clustering	Lab: Advanced Data Visualization and Reporting
12	18-Apr-19	1-4 PM	Unstructured Data + Sentiment Analysis	
13	25-Apr-19	1-4 PM	Lab: Unstructured Data + Sentiment Analysis	
14	2-May-19	1-4 PM	Advanced Algorithms + Issues and Challenges	Predictive Analysis

MODULES GOALS

1. Increase ability to discuss data science/A.I./Machine Learning
2. Increase **knowledge of data science/A.I./machine learning** and its functionality
3. Introduce the **fundamental concepts** of data science/A.I./machine learning
4. Aid the understanding of the **practical aspects** of data science/A.I./M.L.
5. Provide an **initial experience** analyzing datasets
6. Provide **stepping stones** to the next 1200 hours

TANGENTIAL GOALS

These modules are only **partially** about

- how to program computers to analyze data
- the entire gamut of data mining techniques
- current and emerging technologies and software (R / Python) used to analyze data

If you know R and Python already – great! But it is not essential. Programming is approached by **osmosis** (through notebooks…)

Corollary: what programming framework will you be using?

PEDAGOGICAL APPROACH

In general, modules are split between **theoretical lectures**, **practical exercises** (R and/or Python notebooks), **group discussions**, and **case studies**.

Lectures may be technical (math, but no proofs), non-technical (heuristics, case studies), or a **mixture** of both.

Philosophy: even if you never need to know the details, it's beneficial to have an idea what details **look and smell like**.

Please do not hesitate to ask questions at any moment.

TRAINING CHALLENGES

No examinations, no homework… **no motivation**?

School days may be (happily) **far removed**

Survey course → some **crucial details are not covered**

Heterogeneous groups: reasons, skills, interests, experience

Hype and **expectations**

Real-world concerns (you) vs. **Theoretical** concerns (DAL)

"The early stages of education *have* to include a lot of lies-to-children, because early explanations have to be simple. However, we live in a complex world, and lies-to-children must **eventually be replaced** by more complex stories if they are not to become delayed-action genuine lies."

(T. Pratchett, I. Stewart, J. Cohen, *The Science of Discworld*)

lies-to-children: "as much as they can understand"
lies-to-bosses: "as much as they need to know"
lies-to-patients: "as much as is required to keep them from worrying"
lies-to-ourselves: ...

Mr. Gustave: I'm not angry with Serge. You can't blame someone for their basic lack of moral fiber. [...] It's not his fault, is it?

Mustapha Zero: I don't know. It depends.

Mr. Gustave: Well, you can say that about most anything. "It depends." Of course, it depends.

Mustapha Zero: Of course it depends. Of course it depends.

Mr. Gustave: Yes, I suppose you're right. Of course it depends.

(*The Grand Budapest Hotel*, 2014)