PlusPlus Academy

Data Science Overview

Data Science

An Bình

7/2022

Table of Contents

What is Data Science

1 Data Science In Real Life

DS là việc tận dụng các kỹ thuật về thống kê (**statistical**) và công cụ tính toán (**computational**) lên data để mang lại giá trị có ý nghĩa cho việc phát triển doanh nghiệp, hoặc để giải quyết một bài toán nào đó trong thực tế (**real world**).

What is Data Science

1 Data Science In Real Life

- Statistical: Phân tích, thống kê, thuật toán ML, ...
- Computational: Code, tận dụng khả năng tính toán của máy tính để thực hiện statistical
- Real World: Các bài toán doanh nghiệp gặp phải trong thực tế, không phải các competitions, ...

What is Data Science

1 Data Science In Real Life

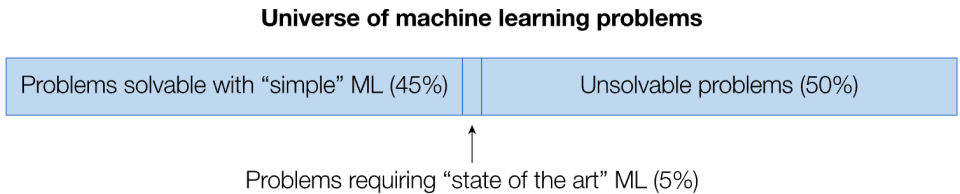
Data science =

- statistics
- data collection
- data preprocessing
- machine learning
- visualization
- business insights
- ...

What is Data Science NOT

1 Data Science In Real Life

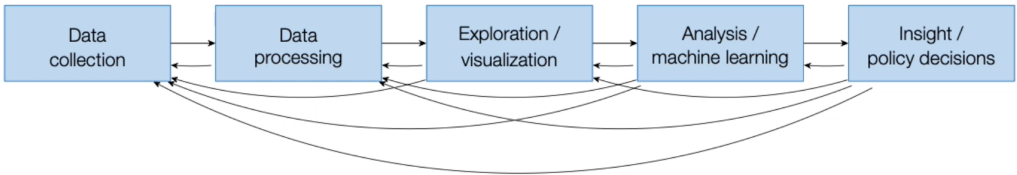
- Data science không phải (chỉ) machine learning.



- Data science không phải (chỉ) statistic và big data.

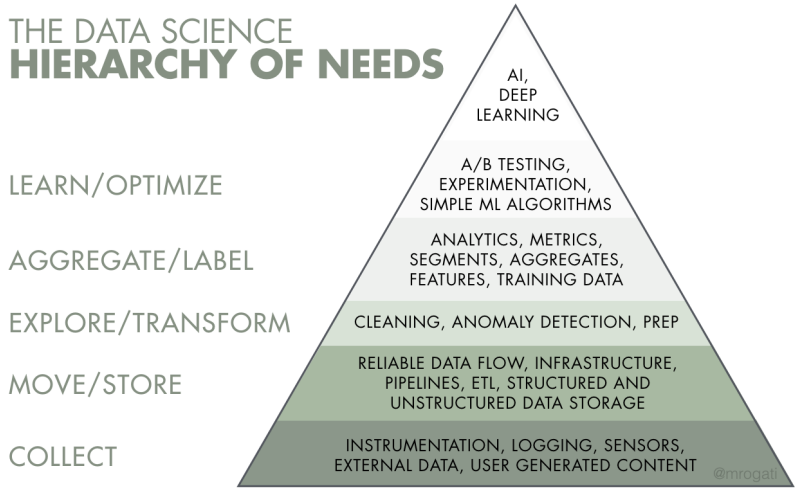
Back To What Data Science Is

1 Data Science In Real Life



What Data Scientist Do

1 Data Science In Real Life



Data Science Example

1 Data Science In Real Life

Gender word in professor reviews

Gendered Language in Teacher Reviews

I've had trouble keeping this site up continuously during COVID. As of March 2021, I'm now trying a new strategy to cache common queries on the server even when the underlying database is down. If you find that many searches don't change the results, that's why.

This interactive chart lets you explore the words used to describe male and female teachers in about 14 million reviews from RateMyProfessor.com.

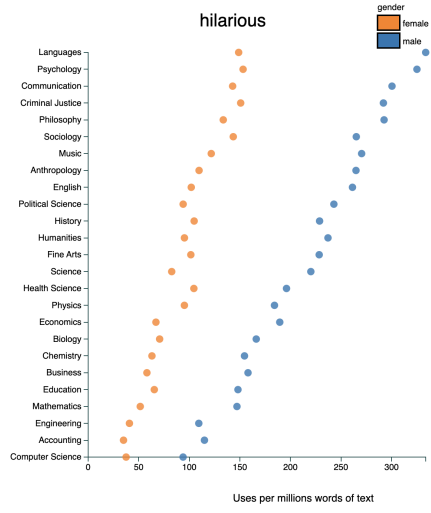
Not all words have gender splits, but a surprising number do. Even things like pronouns are used quite differently by gender.

Search term(s) (case-insensitive): use commas to aggregate multiple terms

hilarious

All ratings | Only positive | Only negative

You can enter any other word (or two-word phrase) into the box above to see how it is split across gender and discipline: the x-axis gives how many times your term is used per million words of text (normalized against



Data Science Example

1 Data Science In Real Life

Aspect Based Sentiment Analysis

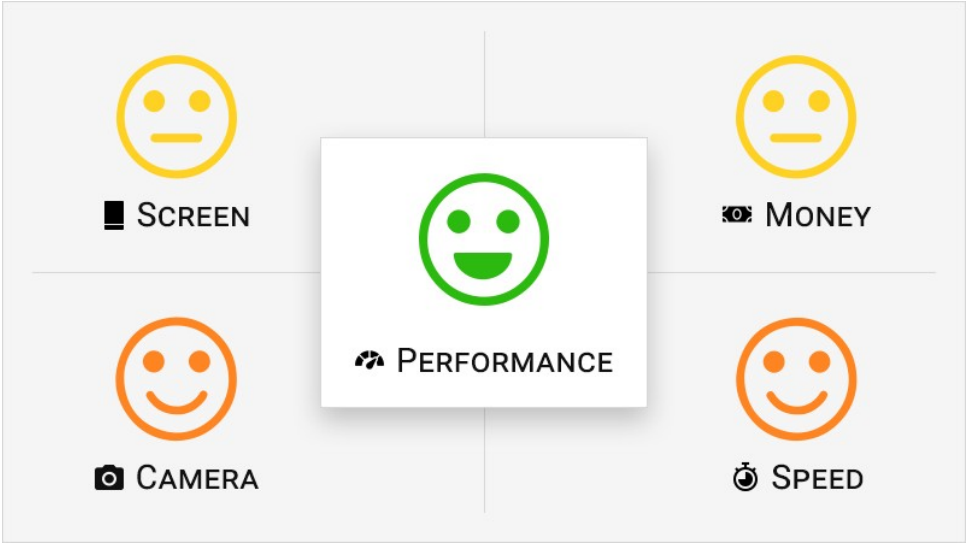


Table of Contents

Course Objective

2 Outline

Một số mục tiêu của khóa học:

- Hiểu được data science pipeline.
- Biết sử dụng một số tools để thực hiện các phần trong pipeline trên.
- Biết kỹ thuật data collection.
- Biết data processing.
- Biết sử dụng một số thuật toán ML/DL vào các dạng bài toán thích hợp.

Topic Covered

2 Outline

Một số topics trong khóa học (có thể thay đổi theo trend):

- **Data Collection and management**: Structure data, metrics and vectors, free text processing, etc
- **ML and DL**: Model selection, metrics evaluation, linear, non-linear, clustering, boosting, recommendation system, etc

Table of Contents

Conda Environment

3 Environment Setup

Cài đặt miniconda

macOS installers

macOS

Python version	Name	Size	SHA256 hash
Python 3.9	Miniconda3 macOS Intel x86 64-bit bash	56.0 MiB	007bae6f18dc7b6f2ca6209b5a0c9bd2f283154152f82becf787aac709a51633
	Miniconda3 macOS Intel x86 64-bit pkg	62.7 MiB	cb56184637711685b08f6eba9532cef6985ed7007b38e789613d5dd3f94ccc6b
	Miniconda3 macOS Apple M1 ARM 64-bit bash	52.2 MiB	4bd112168cc33f8a4a60d3ef7e72b52a85972d588cd065be803eb21d73b625ef
	Miniconda3 macOS Apple M1 ARM 64-bit pkg	63.5 MiB	0cb5165ca751e827d91a4ae6823bfda24d22c398a0b3b01213e57377a2c54226
Python 3.8	Miniconda3 macOS Intel x86 64-bit bash	56.4 MiB	f930f5b1c85e509ebbf9f28e13c697a082581f21472dc5360c41905d10802c7b
	Miniconda3 macOS Intel x86 64-bit pkg	63.1 MiB	62eda1322b971d43409e5dde8dc0fd7bfe799d18a49fb2d8d6ad1f6833448f5c
	Miniconda3 macOS Apple M1 ARM 64-bit bash	52.5 MiB	13b992328ef088a49a685ae84461f132f8719bf0cabc43792fc9009b0421f611
	Miniconda3 macOS Apple M1 ARM 64-bit pkg	63.8 MiB	e92fd40710f7123d9e1b2d44f71e7b2101e3397049b87807ccf612c964beef35
Python 3.7	Miniconda3 macOS Intel x86 64-bit bash	66.0 MiB	323179e4873e291f07db041f3d968da2ffc102dcf709915b48a253914d981868
	Miniconda3 macOS Intel x86 64-bit pkg	72.7 MiB	9278875a235ef625d581c63b46129b27373c3cf5516d36250a1a3640978280cd

Conda Environment

3 Environment Setup

Tạo conda environment với python version 3.9 và env name plusplus:

```
conda create -n plusplus python=3.9
```

Activate environment vừa tạo:

```
conda activate plusplus
```

Cài đặt python package sử dụng pip, ví dụ pandas, numpy:

```
pip install pandas numpy
```

Cài đặt Code editor: Visual Studio Code
Cài đặt jupyter notebook

```
pip install notebook
```

Tạo notebook (thao tác bằng browser)

```
jupyter notebook
```

Mở browser: localhost:8888

Table of Contents

Help and Documentation in IPython

4 Python Recap

Accessing Documentation with ?

```
In  [3]:    L=[1,2,3]
In  [4]:    L.insert?
Type:   builtin_function_or_method
String form:    <built-in method insert of list object at 0x1024b8ea8>
Docstring:      L.insert(index,object) -- insertobjectbeforeindex
```

Wildcard matching

```
In [10]:    str.*find*?
            str.find
            str.rfind
```

Help and Documentation in IPython

4 Python Recap

Timing Code Execution: %timeit

```
In [8]: %timeit L = [n ** 2 for n in range(1000)]
1000 loops, best of 3: 325  s  per loop
```

Shell Commands in IPython

```
In [1]: !ls
myproject.txt
In [2]: !pwd
/home/binhna/projects/myproject
```

Help and Documentation in IPython

Không thể navigate vào folder bằng !

Shell-Related Magic Commands

```
In [11]: !pwd
/home/binhna/projects/myproject
In [12]: !cd ..
In [13]: !pwd
/home/binhna/projects/myproject
```

Dùng %

```
In [14]: %cd ..
/home/binha/projects
```

Q&A

Thank you for listening!
Your feedback will be highly appreciated!