

Intro to Data Science

Alistair Walsh

Alistair Walsh

Instructor

PhD candidate at The Melbourne
Brain Centre (The Florey Institute)

Lead Instructor Melbourne Data Science Intensive course - GA

Fellow (by Training) of the Australasian College of Health Informatics

BSc. Hon. Cognitive Neuroscience

BSc. Psychology/Psychophysiology

Dip. Laboratory Technology (Biotechnology)



Introduction

What is Data Science?

What is a Data Scientist?



"Data Scientist' is a Data Analyst who lives in California"

"A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician."

Someone who can collect, statistically explore and analyse data in an efficient and reproducible manner... but who can also translate from Dataese to Peoplese. Oh, and something something machine learning.

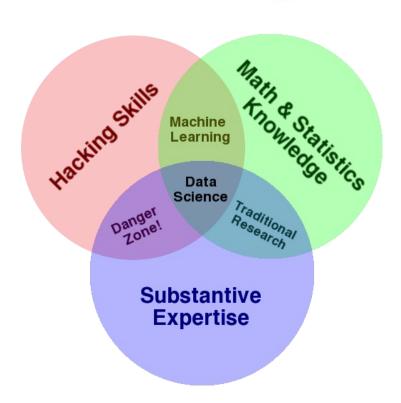
WHAT IS DATA SCIENCE?



A set of tools and techniques for data

Interdisciplinary problem-solving

Application of scientific techniques to practical problems





Data Developer	Developer	Engineer	
Data Researcher	Researcher	Scientist	Statistician
Data Creative	Jack of All Trades	Artist	Hacker
Data Businessperson	Leader	Businessperson	Entrepeneur

Data Science involves a variety of roles, not just one.

Data Science involves a variety of skill sets, not just one.



Business

Product Development

Domain Knowledge

Data Collection

Data Storytelling

ML / Big Data

Structured Data

Unstructured Data

Graph Data

Distributed Data

Parallel Processing Applied Math

Algorithm Design

Linear Algebra

Matrix Calculations

Model Optimization

Dimensionality Reduction Programming

Data Acquisition

Data Cleaning

Object-Oriented Programming

Database Administration

Data Engineering

Natural Language Processing Statistics

Spatial Statistics

Temporal Statistics

Descriptive Statistics

Data Visualization

Feature Selection

Multi-Armed Bandit

Study Design

Model Evaluation

Skills and Self-ID Top Factors Data Businessperson Data Developer Data Researcher Data Creative **Business** ML/Big Data Math/OR Programming Statistics



Roles in Data Science

These roles prioritize different skill sets.

However, all roles involve some part of each skillset.

Where are your strengths and weaknesses?

Why Now?



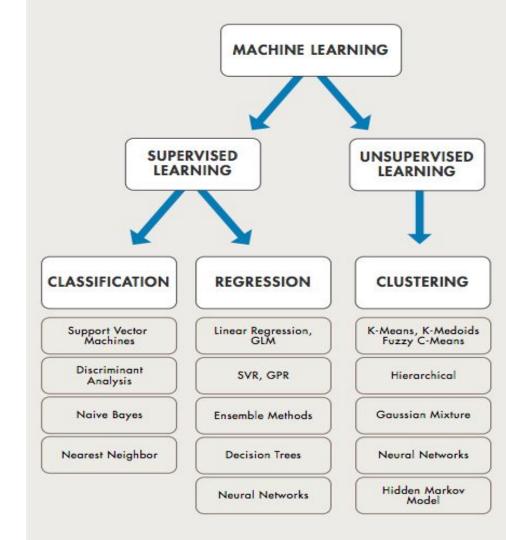
- 'Big Data'
- + Algorithms
- + Computing Power = Data Science

Problems We Solve



Understand	Unsupervised Learning	What has happened?
Predict	Supervised Learning	Given this, what will happen?
Influence	Feature Importance	What factors must I change to bring about what I want

How we solve the problems we solve





The Data Science Workflow

DATA SCIENCE WORKFLOW

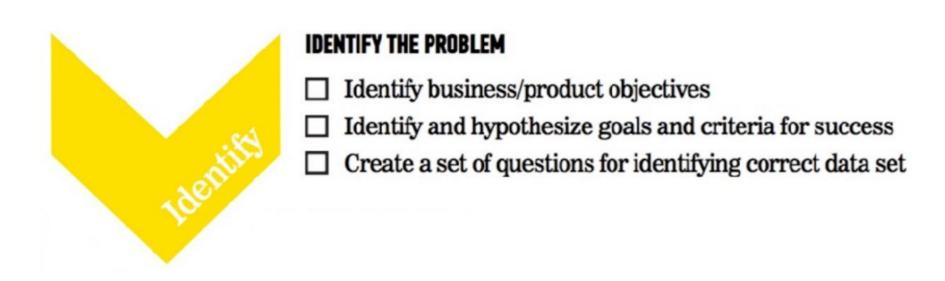




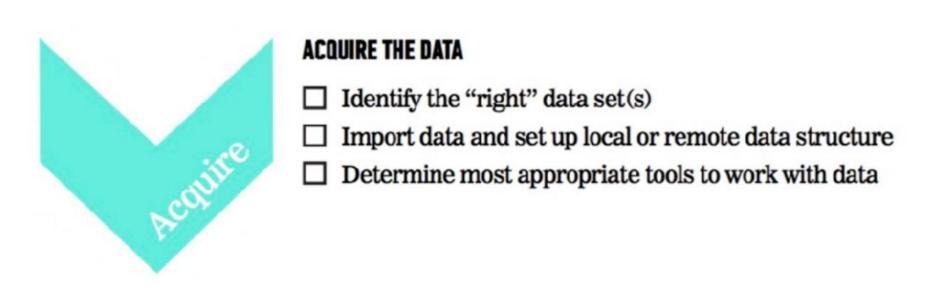
The steps:

- 1. Identify the problem
- 2. Acquire the data
- 3. Parse the data
- 4. Mine the data
- 5. Refine the data
- 6. Build a data model
- 7. Present the results









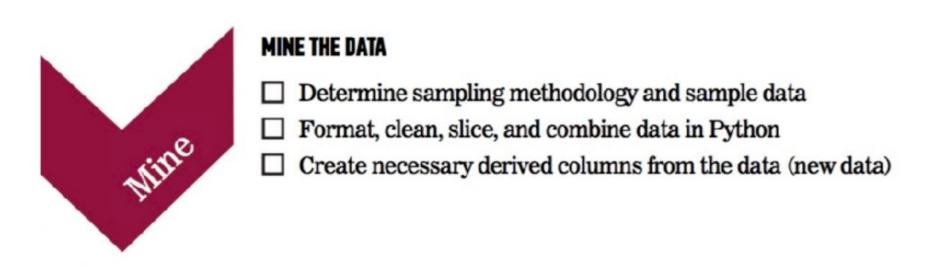




PARSE THE DATA

- Read any documentation provided with the data
- ☐ Perform exploratory data analysis
- ☐ Verify the quality of the data

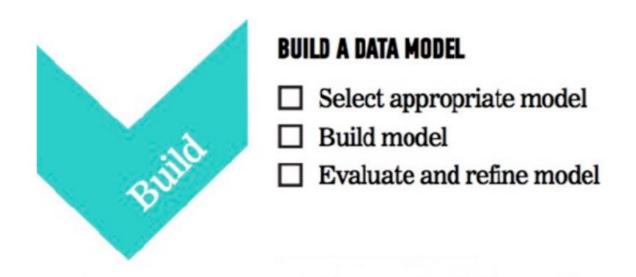
















PRESENT THE RESULTS

- Summarize findings with narrative, storytelling techniques
- ☐ Present limitations and assumptions of your analysis
- ☐ Identify follow up problems and questions for future analysis

Data Science Toolbox





SQL







Let's Code!



