Actuaries Institute

What is Data Science?

Anthony Tockar

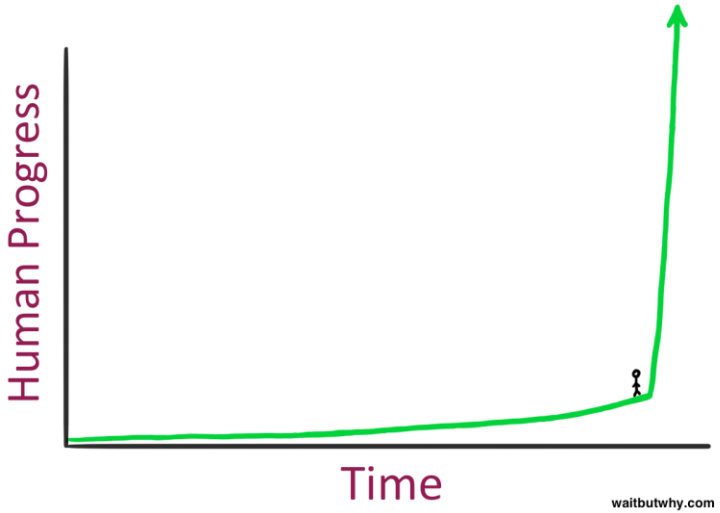
Actuaries Institute

Agenda

- Why Data Science?
- Definitions
- Requirements
- Data Science in Practice
- How to Get Started
- Q&A / Discussion

Why Data Science

"Most people don't understand just how quickly machine intelligence is advancing"
Elon Musk



Why Data Science

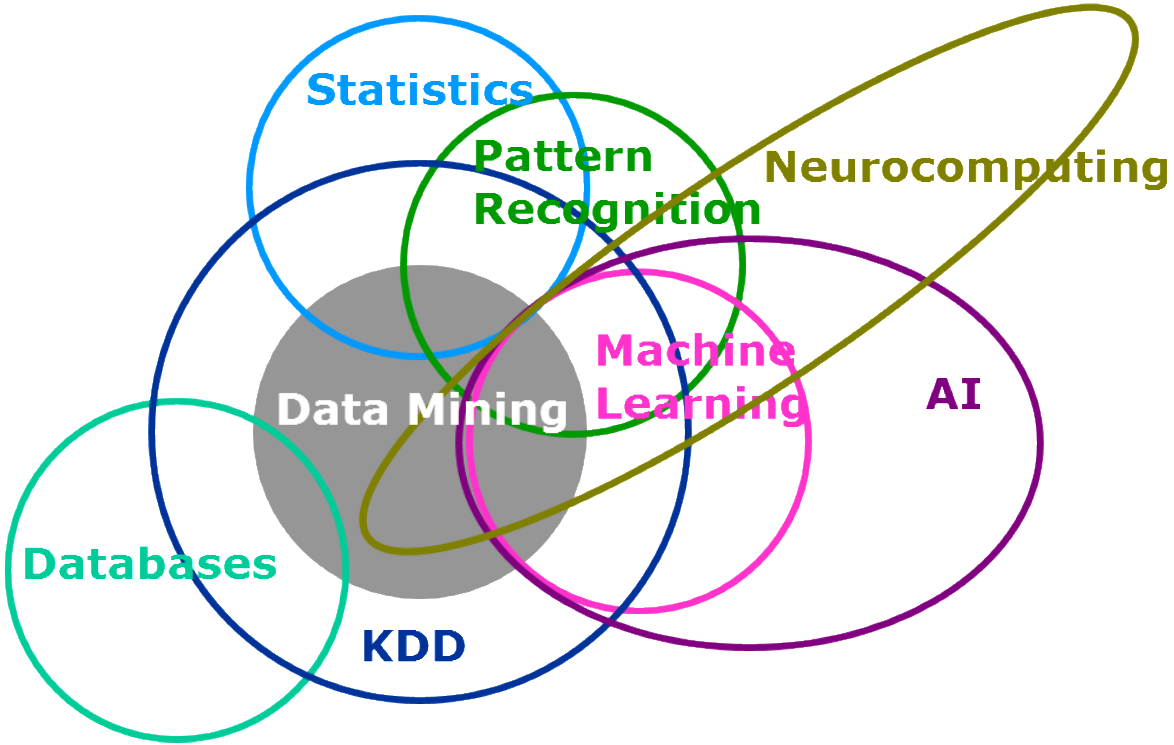
- Businesses need help making decisions in a changing, competitive world

- Data is at the centre of all this - Over the past years we have collected more and more data. We are now in a position to analyse it, to build new products, new industries and new ways of life out of it. Skills in extracting value from data to create change are coming to the forefront

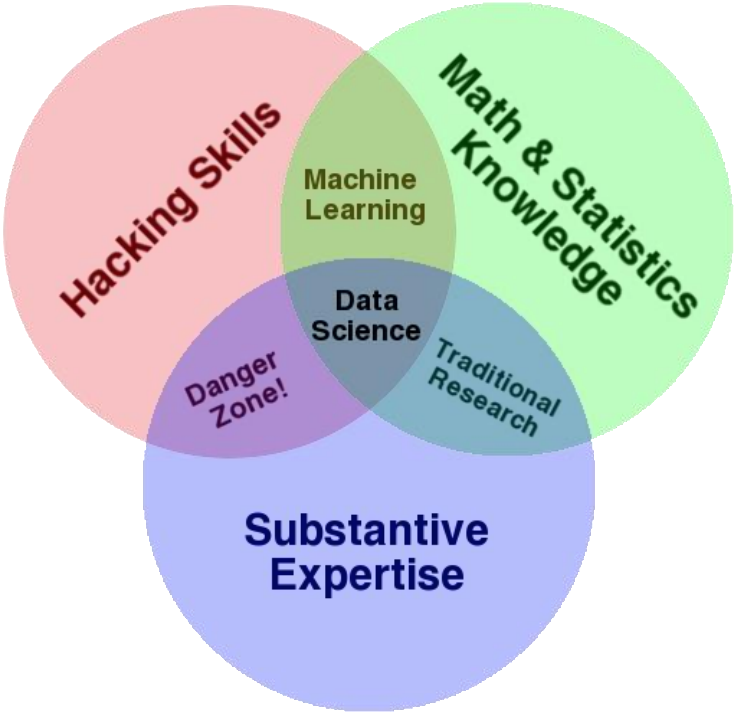
Definitions

Data Science	Extracting valuable, actionable **insights** from **data** to support and enhance **decision-making** / Using both **data** and **science** to create something **new**
Big Data	High-**volume**, high-**velocity** and/or high-**variety** information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision-making and process automation.
Machine Learning	An application that provides systems the ability automatically **learn** and **improve** from experience **without** being explicitly programmed. In business, it is commonly referred to as **predictive analytics**.
Artificial Intelligence	Intelligence exhibited by machines, rather than humans or other animals.

Definitions - interaction



Definitions – Drew Conway's Data Scientist



Core Data Roles

Solutions architect

- Translates requirements created by functional analysts into the architecture for that solution, and describes it through architecture and design artifacts. The rest of the development team then uses those artifacts to implement the solution.

Data engineer

- Development, building data applications and infrastructure, data modelling, machine learning pipelines
- Operationalise data processes

Data analyst

- Describe what is there and what is happening
- Examine data facts and characterisation with retrospective, deductive reasoning

Data scientist

- Investigate and estimate things with proven methods – statistics, inference, prediction, modelling, inductive reasoning
- Apply the scientific method to solve business problems – research, exploration, experimentation, discovery

Mindset

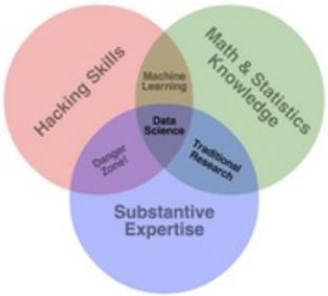
Patil's first flippant answer to "what kind of person are you looking for when you hire a data scientist" was "someone you would start a company with". That's an important insight: we're entering the era of products that are built on data. We don't yet know what those products are, but we do know that the winners will be the people, and the companies, that find those products… No one in the nascent data industry is trying to build the 2012 Nissan Stanza or Office 2015; they're all trying to find new products. In addition to being physicists, mathematicians, programmers, and artists, they're entrepreneurs.

What is Data Science, Mike Loukides

Skills

Programming / Algorithms

- Back / front-end
- Programming languages
- Optimisation
- Simulation
- Computer science



Machine Learning

- Supervised learning
- Unsupervised learning
- Reinforcement learning
- Deep learning
- AI

Statistics

- Statistical methods
- Maths and science
- Surveys
- Time series

Visualisation / Reporting

- Dashboards
- BI
- Design
- Communication

Data Management / Big Data

- Data collection
- Data manipulation
- Data warehousing
- Unstructured data
- Distributed computing

Tools


JavaScript
AWS
NoSQL
Hadoop
Pig
Shark
Bash
Redis
SPSS
Java
R
Spark
JMP
C++
Perl
PHP
Tableau
Cassandra
Python
C
sed
Scala
Cascading
QlikView
awk
Clojure
SAS
Redshift
Fortran
Hive
Ruby
Excel
HBase
Mahout
objective-C
SQL
Matlab
ElasticSearch

Breadth vs Depth

T-shaped professionals: 'T' represents breadth of skills, across the top, with depth in one area represented by the vertical bar. T-shaped professionals can more easily work in interdisciplinary teams than those with less breadth and can be more effective than those without depth. Data science is an inherently collaborative and creative field, where the successful professional can work with database administrators, business people, and others with overlapping skill sets to get data projects completed in innovative ways.

Analyzing the Analyzers, O'Reilly 2013

DS in Practice - Consulting

- Create decision science framework, knowledge base and training materials
- POC evaluating ML solutions
- Huge data ingestion project in Hadoop
- Customer segmentation and churn prediction

DS in Practice – Other Projects

- Statistical case estimation
- English-to-SQL translator
- Reinforcement learning – Atari Breakout
- Differential privacy
- Age imputation
- Social network visualisation and analysis

Actuaries Institute

How to Get Started

Online

- Data Science Central
- KDnuggets
- Data Science Weekly

Books

- Joel Grus - *Data Science from Scratch*
- Hastie et al. - *Elements of Statistical Learning*
- Tufte - *The Visual Display of Quantitative Information*

PDFs

- Mike Loukides - *What is Data Science?*
- Vaisman et al. - *Analyzing the Analyzers*

Courses

- Coursera - *Data Science*
- UTS - *Master of Data Science and Innovation*
- USyd - *Master of Data Science*

Meetups

- Data Science Breakfast
- Data Co-op!
- Data Science Sydney
- SURF
- Data Visualisation Sydney

Projects

- Kaggle *(Titanic)*
- Minerva Collective
- Make one up

Plug

Data Visualisation Sydney