ECMM443/COM2015
Introduction to Data Science

Week 1: Introduction

Dr Xiaoyang Wang

Department of Computer Science

x.wang7@exeter.ac.uk

Teaching Team

Module Conveners:
Dr Xiaoyang Wang, x.wang7@exeter.ac.uk
Dr Phil Lewis, p.lewis2@exeter.ac.uk

Teaching Assistants:
Zihao Chen, zihao.chen@exeter.ac.uk
Heng Yang, h.yang5@exeter.ac.uk
Tiankuo Hou, th615@exeter.ac.uk

Data Science

University of Exeter

What is data science and what do you expect from this module?

Go to `https://www.mentimeter.com/`, and enter the code **3144 4052**
Or scan the following QR code

Some results from the 2024 cohort are on the next page.

Menti Results

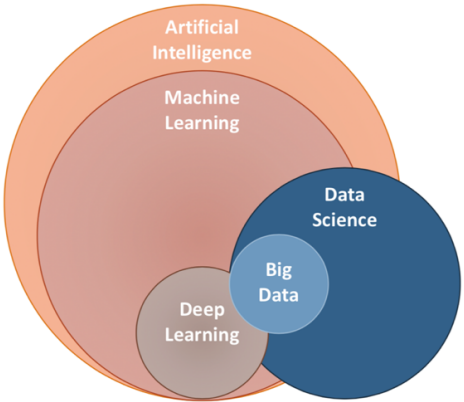
"How would you define data science?"

"What do you expect from this module?"

Need More on Python?

Optional Module: COMM109 Programming with Python - it covers the basics

Artificial Intelligence? Machine Learning?



Venn Diagram for data science, AI and ML

Data Science:

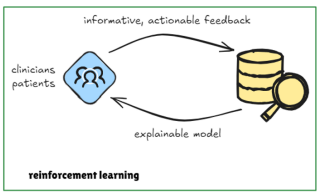
Multidisciplinary
Maths and statistics
Computer science
Domain knowledge

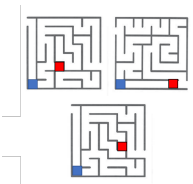
https://gmggroup.org/wp-content/uploads/2019/07/AI-Venn-diagram.png

My Research

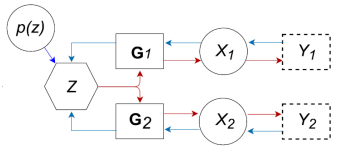
My research focuses on *Machine Perception and Intelligence*
`https://wang-xiaoyang.github.io/`

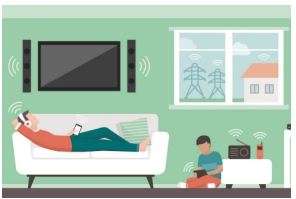
- Reinforcement learning, decision support systems, human-centric AI





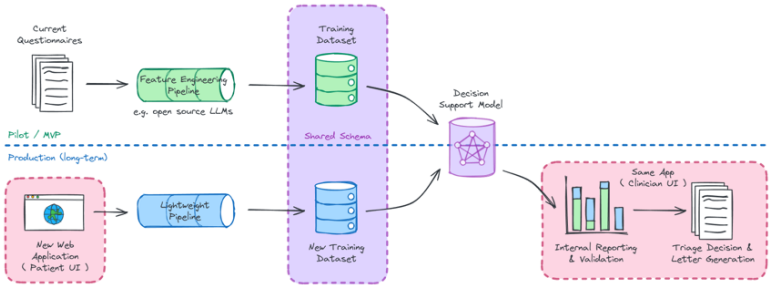
- Computer vision, sensor fusion





https://www.bristol.ac.uk/news/2018/may/opera-project.html

Example: Clinical Decision Support System

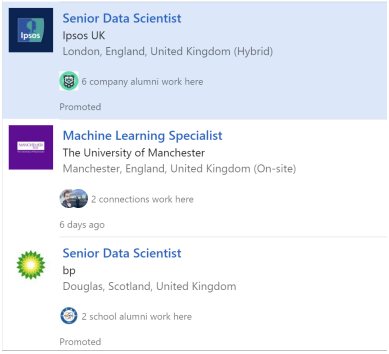


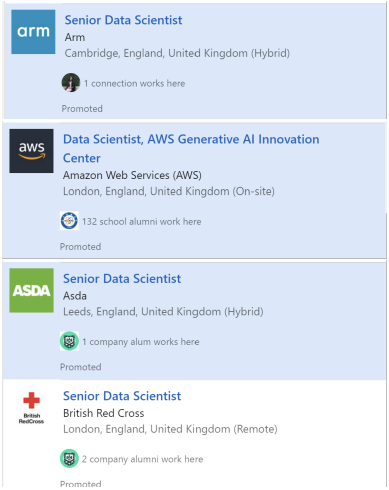
Decision support systems for patients triaging in a virtual ENT clinic

Industries

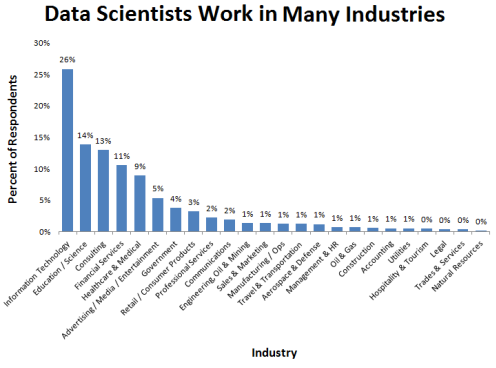
Industries

I did a LinkedIn Job search ...



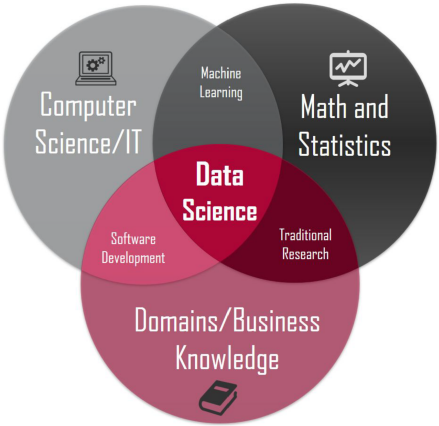


Data Science in Industries



https://businessoverbroadway.com/2016/05/09/industry-differences-in-data-science-roles-skills-and-project-outcomes/

Aspects of Data Science



This figure is from Dr Rudy Arthur's slides

Example: Data for Good

Link: Data for Good, by Meta

"Social Connectedness Index"

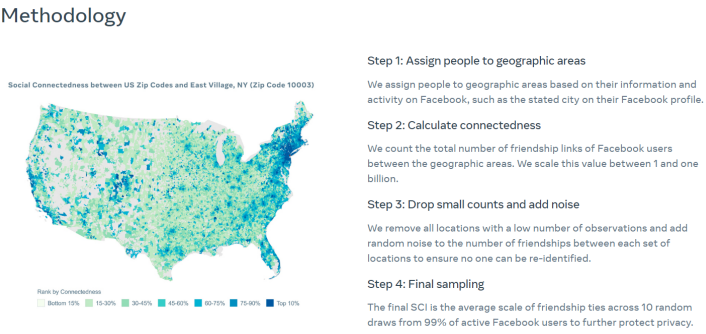
What: "This index measures the strength of connectedness between two geographic areas as represented by Facebook friendship ties."
For what: "SCI can reveal important insights about economics, social mobility, and health."

https://dataforgood.facebook.com/dfg/tools/social-connectedness-index

Example: Data for Good

Link: Data for Good, by Meta
"Social Connectedness Index"

How:



https://dataforgood.facebook.com/dfg/tools/social-connectedness-index

Example: Data for Good

Link: Data for Good, by Meta
"Social Connectedness Index"

Impact: "The SCI should allow researchers to overcome some of the measurement challenges that have held back empirical research on the role of social interactions in finance, economics, and the broader social sciences."

Related publications: `Chetty, R., Jackson, M.O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R.B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M. and Johnston, D., 2022. Social capital I: measurement and associations with economic mobility. Nature, 608(7921), pp.108-121.`

https://dataforgood.facebook.com/dfg/tools/social-connectedness-index

Module Overview

University of Exeter

In this module, you will:

- Learn about the broad field of data science
- Be introduced to some core competencies and applications
- Improve your programming in Python
- Learn basic statistics and practice data analysis

Schedule

Week 1	Introduction, Admin
Week 2	Visualisation
Week 3	Summary Stats
Week 4	Probability
Week 5	Hypothesis
Week 6	
Week 7	Regression
Week 8	Databases
Week 9	Ethics
Week 10	Spatial Data Analysis
Week 11	Networks
Week 12	Text

* subject to minor adjustments

Plus one workshop per week.
TAs will host the workshop and help with any questions you might have.
Please check your **timetable**!

* Please come to talk with us if you have clashes in your timetable...

ELE-Exeter's e-Learning Platform

University of Exeter

For course information, please go to the
ECMM443 ELE Page
COM2015 ELE Page
Forum
Module Information
Recap

Some Tips

Watch the videos - Thanks to Dr. Rudy Arthur!
Come to workshops with questions; Bring your own laptop if that works best for you
For Python environment management, you could try Anaconda
Practice as much as you can
Don't assume data are complete/without anomalies
Always think about why

Assessments

Module information page: Introduction to Data Science - 2024 entry

The assessments are in two parts:

- Coursework - 20%: Data analysis practice using Python; The coursework document will be published at least 4 weeks before the deadline; **Deadline: ∼Week 9**
- Coursework submissions should be **anonymous** - please put your student number, but not your name.
- Exam - 80%