COMP2200/COMP6200 Lecture 1c – What is Data Science?

Greg Baker

23rd February 2026

MACQUARIE
University

MACQUARIE
University

COMP2200/COMP6200 Lecture 1c – What is Data Science?   Greg Baker   1/4

What is data science?

Applying the scientific method to commercial data

But what does that involve?

MACQUARIE
University

COMP2200/COMP6200 Lecture 1c – What is Data Science?   Greg Baker   2/4

Data scientists often work with a lot of data

https://www.recoverytrial.net/

- Trial with 50,000 patients on Covid treatments
- One of the largest medical trials in history
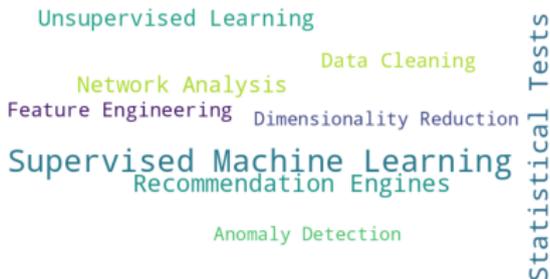

https://www.nature.com/articles/
nature11421

- (2010) Ran a test with 61 million users to encourage voting
- Not even a particularly large data science project

Different scale of data: data scientists often need different skills and tools compared to typical statisticians

MACQUARIE
University

COMP2200/COMP6200 Lecture 1c – What is Data Science?   Greg Baker   3/4

Working with modern analyses

Unsupervised Learning

Data Cleaning

Network Analysis

Feature Engineering

Dimensionality Reduction

Supervised Machine Learning

Recommendation Engines

Anomaly Detection

Statistical Tests

- **Unsupervised Learning** = "break data into groups"
- **Supervised Machine Learning** = "make a predictive model"
- **Statistical tests** = "did the change have any effect?"
- **Network analysis** = "understand relationships and flows"
- **Recommendation engines** = "if you like this, you'll like that"

Data scientists and statisticians only have a small amount of overlap in the algorithms they use