
irds

Administration?

- The course will involve
  - Lectures
  - Keynote Lectures
  - Guest Lectures
  - Mini-Project
  - Literature Review

  - TA Help with the Mini-Project and Literature Review.
  - No lab sessions.

Next Mon, Thu

- Ethics in Data Science
- PrivApprox: Privacy-preserving Stream Analytics, Pramod Bhatotia in IFG07

Choosing Mini-Projects

- There are default datasets and projects and lit review areas.
- BUT
- More interesting if it matches your future interests.
- Talk to staff – do they have interesting data to analyse: you need to anyway to determine a supervisor
- Criterion: only “real” data is allowed. Severely preprocessed data for e.g. benchmarking does not wash.
  - No imagenet, cifar, etc. etc. for an ML project.

Subjects of interest…

What is IRDS for you?

Problem Understanding and Planning

- What do people want?
- What do people really want?
- What do people have?
- What do people really have?
- Evaluation. Evaluation. Evaluation.
- Is it possible? Is it feasible?
- How much data do I need?
- Collaborate, don’t apply.

Furniture Store Advertising Example

What do people want?

- As someone describes: listen carefully!
  - Language games
  - They will try to tell you what they think you want to know. They are usually wrong about what you want to know.
  - They will be vague about things you are used to being precise about and vice versa.
- What do you want to know?
  - What is the actual problem they care about, precisely.
  - How do they currently measure that; how might they measure that: are there usable proxies.

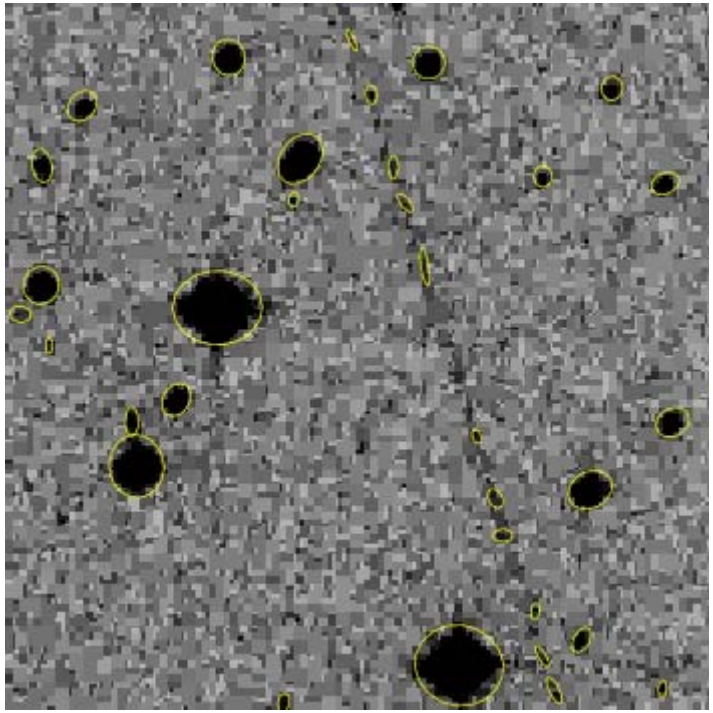

FORA.tv
Don Watson
February 27, 2010
Perth, Australia
Courtesy of Australian
Broadcasting Corporation

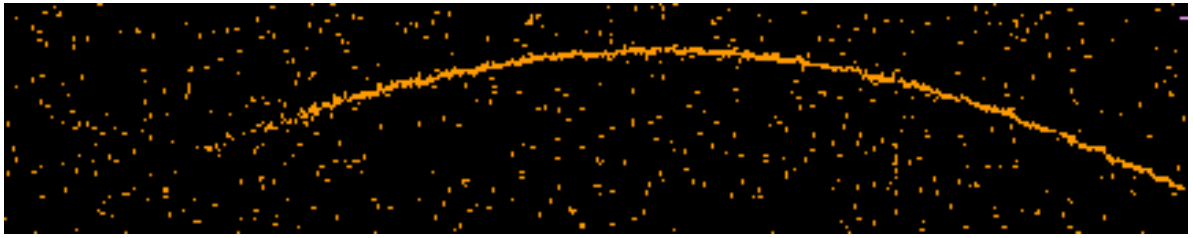
Cutting through

- Ask for specific examples. Not just one. Three.
- Make explicit translation tables: Maybe where you say “parameter”, we would say “variable”?
- Use clear English. Techno-speak is as bad as management speak.
- IDK why computer scientists insist on using acronyms? Don’t.

What do they have?

- Stay sceptical?
- Do they really have it?
- Is it what they say it is?
- Is there missing data?
- Example





Evaluation

- Constantly think about evaluation.

- (How) do they currently evaluate things?
  - Is that computable?
  - If not, is there a computable proxy to that evaluation?
  - Is there a good/better way to evaluate things?
  - In ML, always start with a supervised problem.

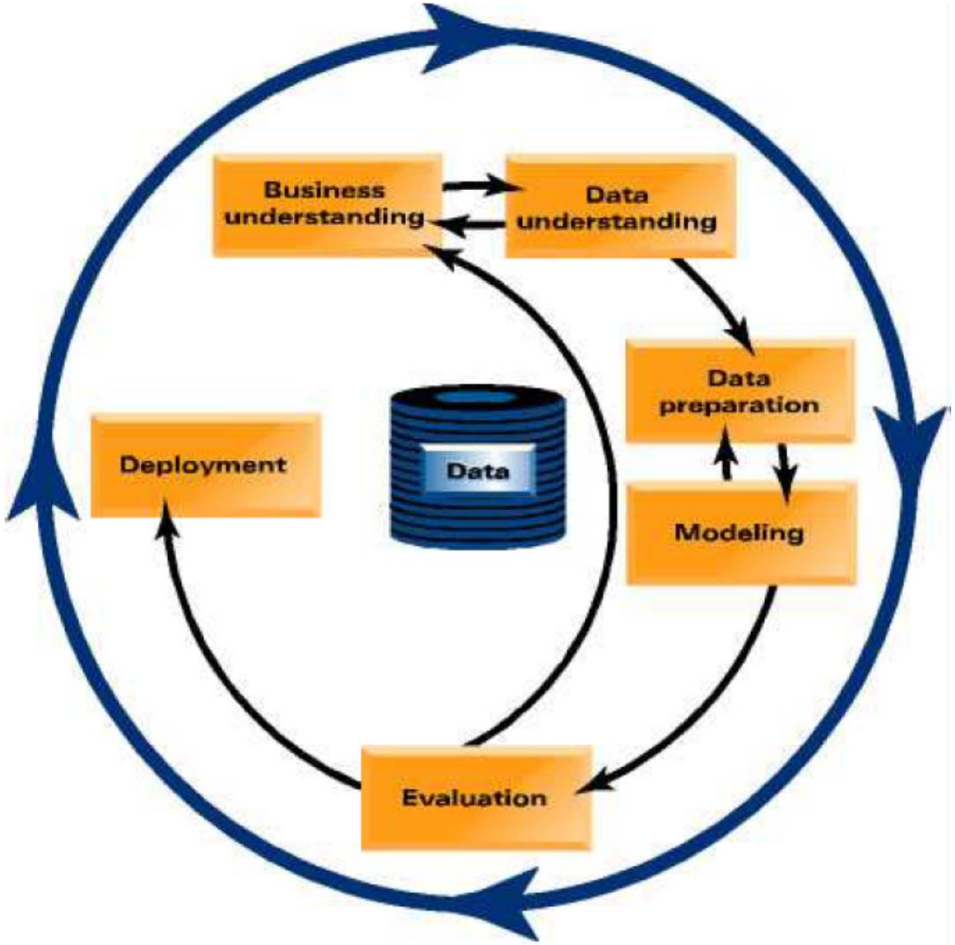
Is it (im)possible?

- This is an important point of exploration
- Example:
  - You are trying to predict the performance capability of a racing car given fixed measurements about the car.
  - The car races in multiple races in a season before the car component measurements are tweaked.
  - How do you assess the whether you might not be able to do this?
- Think about signal to noise ratios.
- Also what are the costs?
  - Data cleaning, collection etc.

How much data? How long will it take?

Collaborate, don’t apply

- CRISP Cross-industry standard process.



Plan for all parts
Plan to iterate
Plan to communicate
Plan to integrate

Summary

- What do people want?
- What do people really want?
- What do people have?
- What do people really have?
- Evaluation. Evaluation. Evaluation.
- Is it possible? Is it feasible?
- How much data do I need?
- Collaborate, don’t apply.