Thoughts on Machine Learning

Chris Mainey

Patient Safety Lead - System Analysis & Delivery, NHSE/I

c.mainey@nhs.net

mainard.co.uk

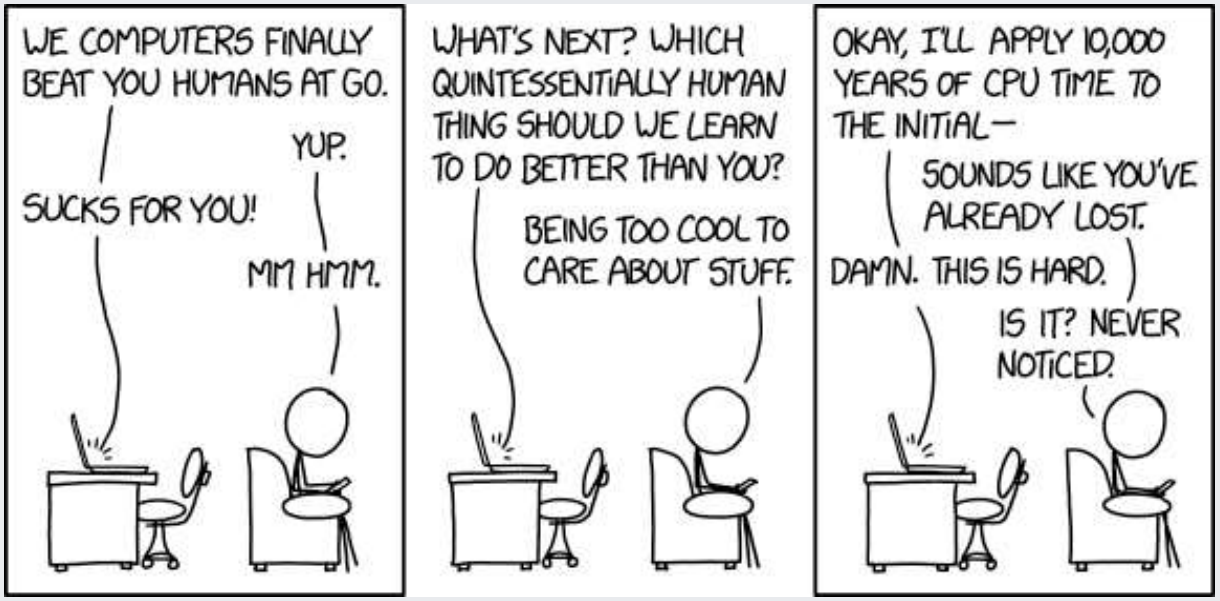
github.com/chrismainey

twitter.com/chrismainey

What is machine learning?

> Machine-learning algorithms use statistics to find patterns in massive amounts of data. And data, here, encompasses a lot of things—numbers, words, images, clicks, what have you. If it can be digitally stored, it can be fed into a machine-learning algorithm.

- MIT review



What is machine learning, really?

- Is it statistics?

Yes, sort of.

- Is it AI?

Yes, sort of.

> “There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models.”
>
> - Leo Breiman - Statistical modeling: The two cultures. Statist. Sci. 16 199–215.
>
> "Conflation between explanation and prediction is common, yet the distinction must be understood for progressing scientific knowledge"
>
> - Galit Shmueli - Statistical Science, 2010, Vol. 25, No. 3, 289–310

I'd argue that most applications of 'AI' could equally be described as ML or statistics

Where is it being used?

All over the place

- Prediction and forecasting
- Search engines
- Image analysis

Some recent uses in NHS:

- Reading imaging data, such as mammography scans - EMRAD
- CogStack - Kings & South London and Maudsley
- Predicting hospital attendance /non-attendance - UCLH / UCL
- Analysis of patient experience data - Notts Healthcare

My opinioned framework:

Yes, that's a little pretentious...

- What is your question?
- What's in there?
- Right-tool for the right job
- Is it any good?
- Letting people know

...with and over reliance on XKCD..._ www.xkcd.com

What is your question?

Manager: Can you pull me some data on it please?

Statistician: _...erm, maybe. What's your question.

Manager: Has anything changed in the last year?

Statistician: By 'anything,' what do you mean? I'm happy to have a look at it?

Manager: Great, can you predict next year?

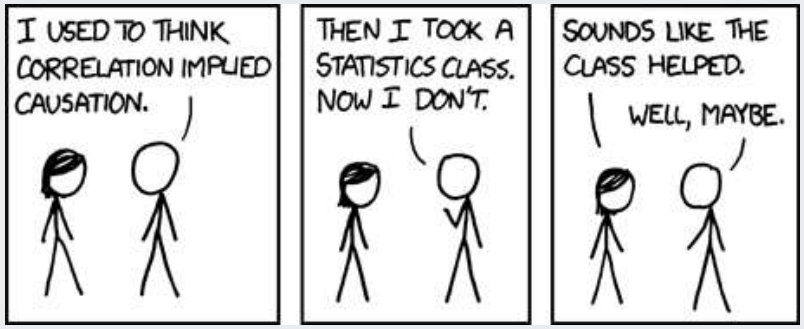
Statistician: Maybe, but what is the thing in question?

Manager: Well, I think you should check all sorts of things. Might be things we don't know about.

Statistician: Pretty hard to answer if it's not clear.

Manager: I don't see why it's so hard. Can't you just check and see if you can find anything?

Statistician: ...erm, again, what is your question?



Important to know your target.

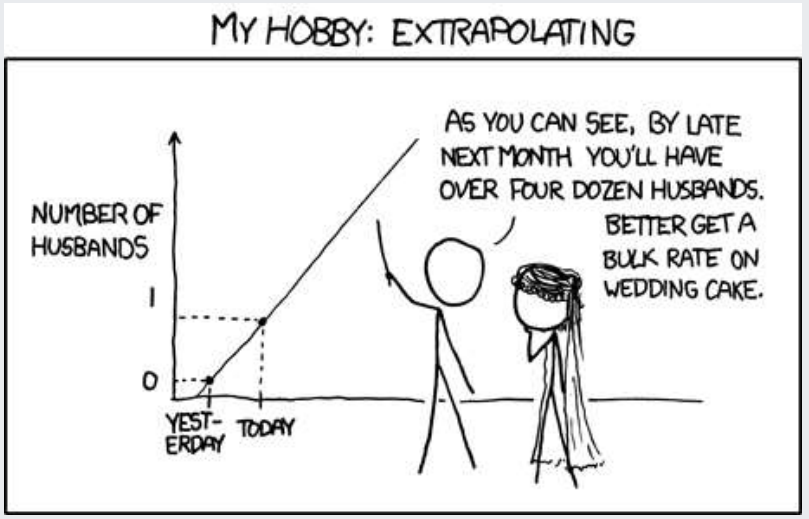
Are you:

- Finding patterns?
- Identifying similar data items?
- Looking for distributions?
- Wanting to know about correlations?
- Predicting events based on attributes, time etc.
- Reduce complexity

What's in there?

Is your dataset suitable for the task at hand?

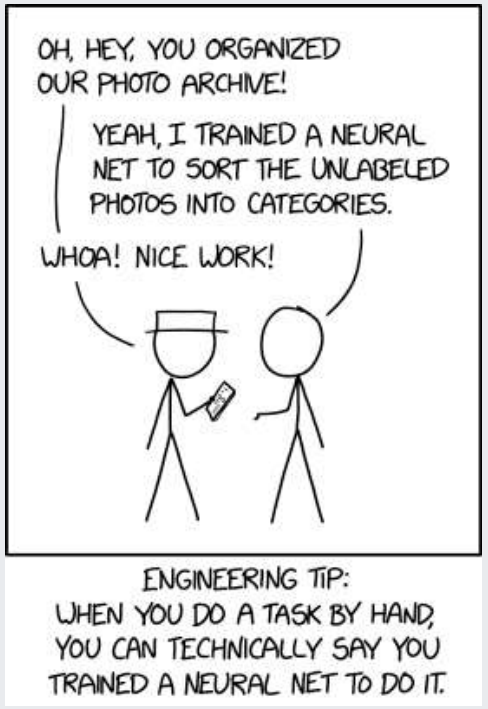
- What is the 'event'?
  - How well represented is it in the data?
  - Do you have enough?
- 'Missingness'
- Coverage
- Correlations/relationships in data



/

Right-tool for the right job?

- Does your model do what you think it does?
- What are the assumptions behind it?
- Does it need to be deep learning?
- Does a simpler model work better?
- Do you have enough events for it?

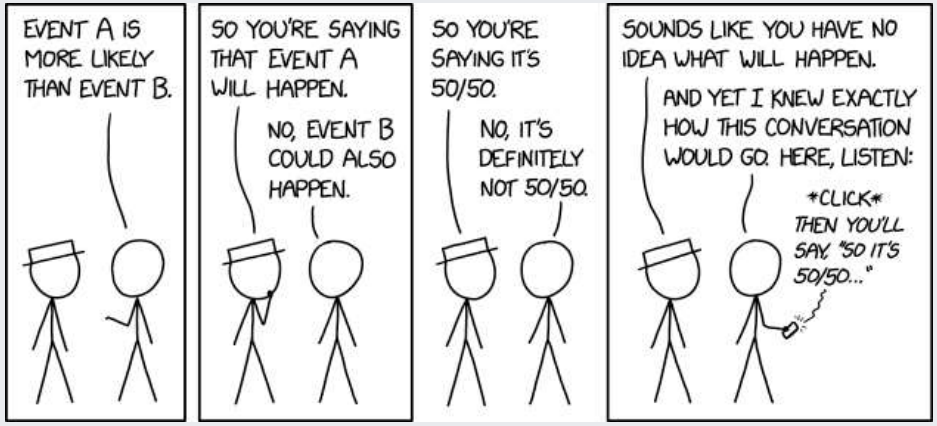


Is it any good?

- Does it work?
- What did you train it on?
- How did you check:
  - Metric?
  - Testing set?
  - A proper test set, not just a validation set / random split
  - Cross-validation / bootstrapping if nothing else
- How does someone else apply it?
- Is it reproducible?
- **BIAS!**

Letting people know

- There are not many of us want the technical detail, but some do
- Who is your audience?
    - How will they understand and use it?
    - How do you convey this to the patients and the public?
- How would you publicise it?
- How does it get put into practise?



Summary

ML is cool.

...but lets do it well!

What is your question?

What's in there?

Right-tool for the right job?

Is it any good?

Letting people know