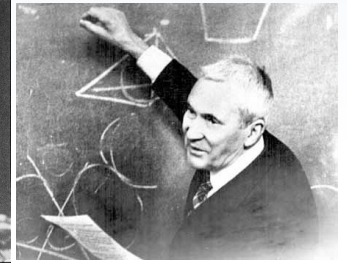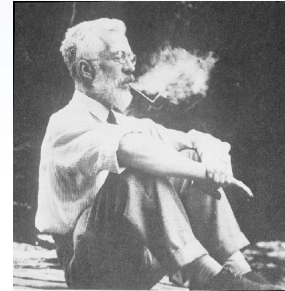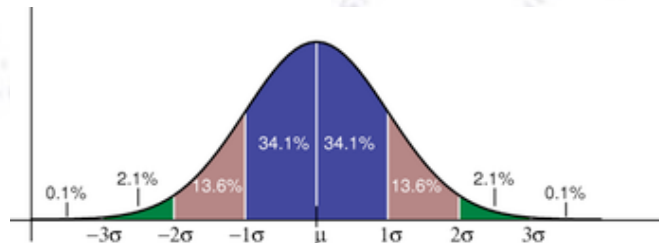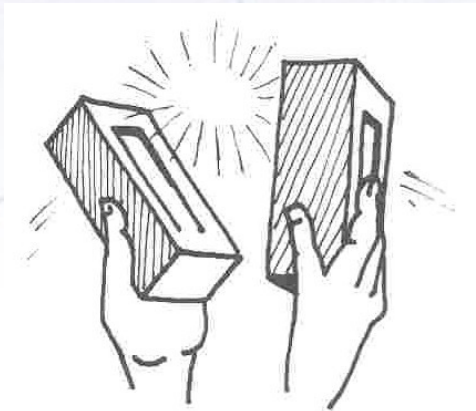# Applied ML

ML cases and industry ideas

Troels C. Petersen (NBI)

"Statistics is merely a quantisation of common sense - Machine Learning is a sharpening of it!"

A great early example

A company producing bricks considered using Machine Learning in the quality control of the bricks. Until now, this had been done manually, with workers **discarding about 4-7% of bricks**.

Based on color, surface and strength of the bricks, a very basic algorithm was trained/optimised and put in place to do the quality control. This worked reasonably well, but unlike the workers, the machine was **discarding 2-30% of bricks!**
WHY?!?
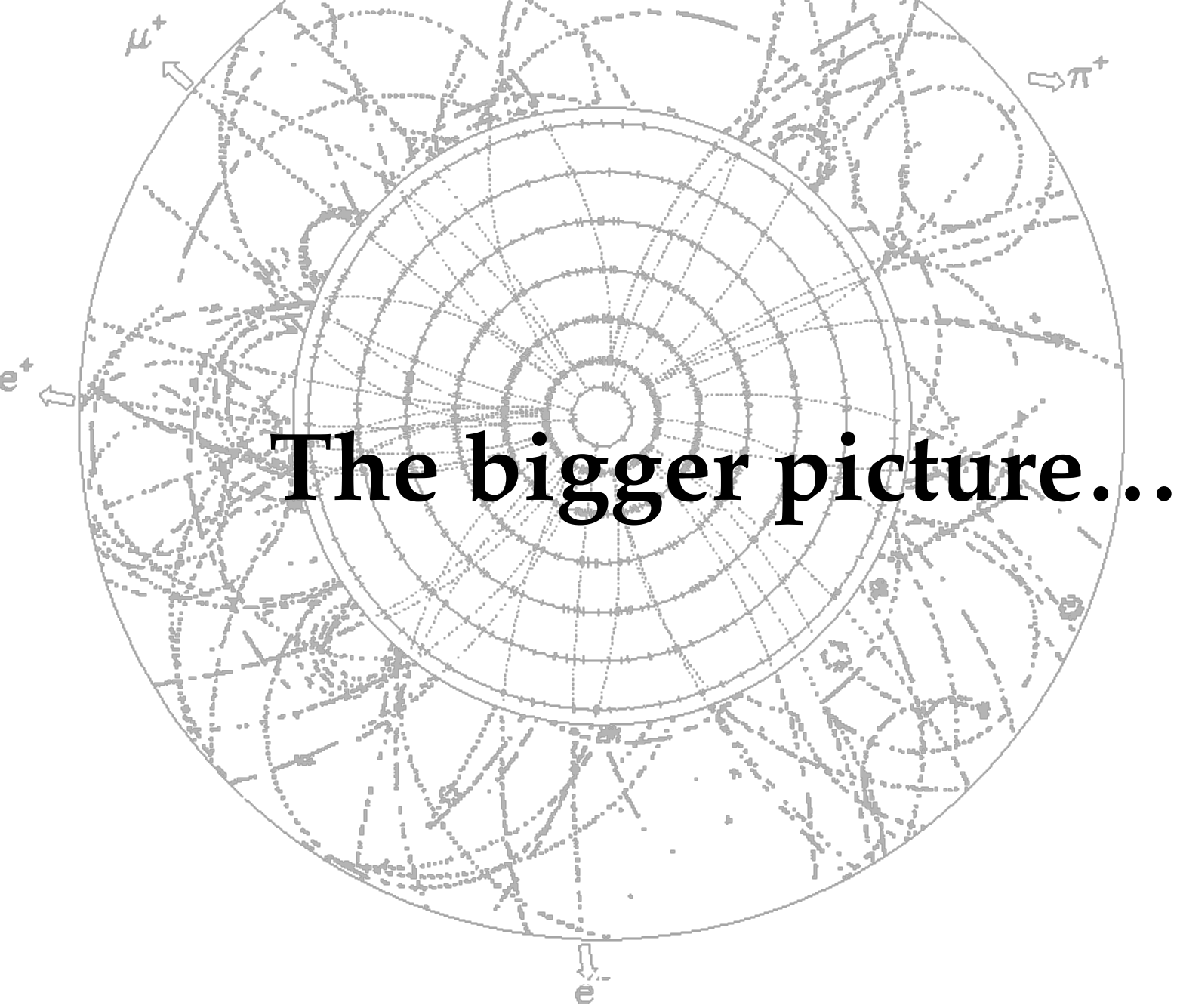
The bigger picture…

Applications

The applications are endless, but it is not easy to see, which cases are clear cut for Big Data Analysis.

Typically, it requires that a few things are in place:
- The business/administration should have much data and many "known cases".
- While ML involves a creative process, it typically improves on repeated actions.
- The more algorithmic/data heavy the business the better.

In light of these points, a business/administration should consider, if they are "ripe" for looking into ML as an option. I suspect that many probably hold back for lack of ideas, knowledge, fear of (over-)spending, etc.
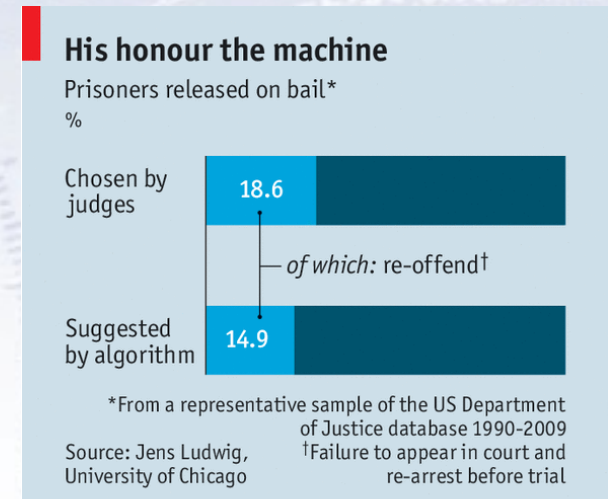
Personal note:
I can (honestly) tell companies, that I know quite a few brilliant students, who not only love to analyse data, but also excel in doing so. Over the years, it has led to quite a few student jobs, for taking a first "cheap" look at the ML options.

Examples:

Who should be released on bail:

In an attempt to find out, who should be released on bail, a group at University of Chicago (CrimeLab) looked into the data of setting bail.

They could decrease the fraction of re-offenders from **18.6**% to **14.9**%. This corresponds (they claim) to what would require 20.000 police officers (2.6 billion $).

His honour the machine
Prisoners released on bail*
%

Chosen by judges — 18.6
— *of which:* re-offend†
Suggested by algorithm — 14.9

*From a representative sample of the US Department of Justice database 1990-2009
Source: Jens Ludwig, University of Chicago
†Failure to appear in court and re-arrest before trial

Economist.com

Heart attack predictions (4 hours in advance):

A study suggested that succes rate would go from 30% to 80% with ML.

Flagging "at risk officers" in US police force:

Using current data, and increase in correct predictions would increase by 12%, while the number of wrong predictions would be reduced by 1/3. (Cop pulling gun at pool).

Chicago is trying to predict, which children have high levels of lead in their blood.

India does actually predict, when it is the best time to sow, and tell farmers.

Ideas:

Turning the Machine Learning idea with myself and (similar minded) tech friends, I've thought of the following ideas (list not exhaustive):

SKAT (ministry of taxation):
SKAT has very large amounts of data on all Danish tax payers going back in time. With **supervised learning** (i.e. based on experience from known cases), it would be very interesting to see, if ML could provide SKAT with a list of suspects, to be investigated further!
Also with **unsupervised learning**, one could divide tax payers into categories, and see if those "alike" (as defined by the clustering) a fraud are also themselves!

Doctors reports:
Doctors have to write a report for every patient visited, and they are typically very standard.
Could one from recording speech (and possibly camera, accelerometer, etc.), create an outline of the necessary report, of course to be checked by the doctor?

Transparency: For transparent answers, use associated Fisher to explain factors. Furthermore, SHAP values allow one to look inside "black box".

What to expect?

Typically, businesses are already good at what they are doing (or they would not be in business anymore!), so the improvements one can expect are typically not that large. A study looked into this, by considering 179 businesses:

Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?

Erik Brynjolfsson
Massachusetts Institute of Technology (MIT) - Sloan School of Management; National Bureau of Economic Research (NBER)

Lorin M. Hitt
University of Pennsylvania - Operations & Information Management Department

Heekyung Hellen Kim
MIT - Sloan School of Management

April 22, 2011

[Ref: SSRN-id1819486]

The study found, that there was a significant improvement going data-driven, that it was not due to reverse causality, and that the general level was…

5-6%