

What Could Possibly Go Wrong?

A (mostly) autobiographic account of things that could go wrong in the life of a data scientist fighting for survival in the corporate world.

“It ain’t what you don’t know that gets you into trouble. It’s what you know for sure that just ain’t so.”

Mark Twain

The Life of a Data Scientist



Understand
the
problem

Know what
you are
doing

Quickly
solve the
problem

Deliver the
solution
and be the
hero

Understanding the Problem

Pedro, the Board wants to know the top 5 customers that we should target.



Understanding the Problem

What 5 customers should we target?

Customer	Expected Income	Probability of Sale
Frodo	€1,072	0.9
Boromir	€513	0.8
Legolas	€155	0.7
Gimli	€75	0.6
Aragorn	€306	0.5
Pippin	€388	0.4
Meriadoc	€827	0.3
Gandalf	€2,495	0.2
Sam	€10,998	0.1

This is where the audience enthusiastically participates.

Understanding the Problem

Choosing Frodo, Boromir, Legolas, Gimli and Aragorn:

- Average probability of sale: 0.7
- Expected income: €2,121
- Probability weighted income: €1,681.7

Choosing Sam, Gandalf, Frodo, Boromir and Meriadoc:

- Average probability of sale: 0.46
- Expected income: €15,905
- Probability weighted income: €3,222.1

Customer	Expected Income	Probability of Sale	Probability Weighted Income
Frodo	€1,072	0.9	€964.8
Boromir	€513	0.8	€410.4
Legolas	€155	0.7	€108.5
Gimli	€75	0.6	€45
Aragorn	€306	0.5	€153
Pippin	€388	0.4	€155.2
Meriadoc	€827	0.3	€248.1
Gandalf	€2,495	0.2	€499
Sam	€10,998	0.1	€1,099.8

Understanding the Problem

- Is there a right answer?
- Did we really understand the problem?
- Why can stakeholders never give us enough context?

It is important to know what use will be given to our output.

- What is the big picture?
- What are we trying to achieve?
- What if we find another way to do it?

Know what you are doing

10. You *won't* need to be a machine learning expert, you *will* need to be an excellent quant and an excellent programmer

J.P. Morgan says the skillset for the role of data scientists is virtually the same as for any other quantitative researchers. Existing buy side and sell side quants with backgrounds in computer science, statistics, maths, financial engineering, econometrics and natural sciences *should* therefore be able to reinvent themselves. Expertise in quantitative trading strategies will be the crucial skill. "It is much easier for a quant researcher to change the format/size of a dataset, and employ better statistical and Machine Learning tools, than for an IT expert, silicon valley entrepreneur, or academic to learn how to design a viable trading strategy," say Kolanovic and Krishnamacharc.

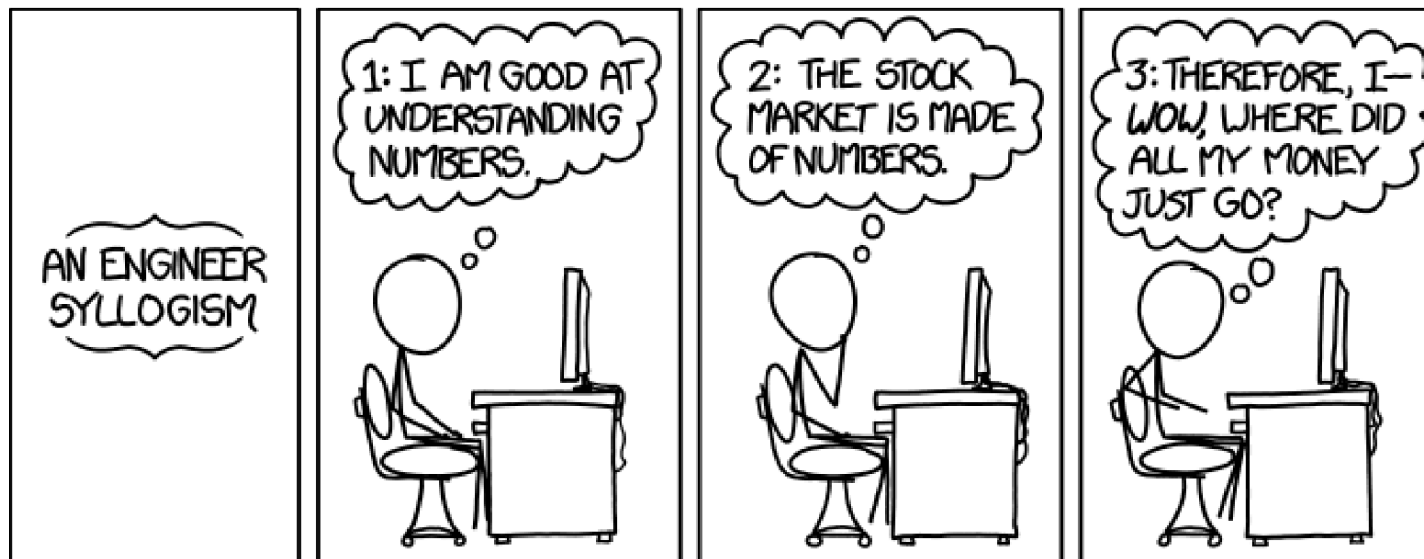
By comparison, J.P. Morgan notes that you *won't* need to know about machine learning in any great detail. – Most of the Machine Learning methods are already coded (e.g. in R): you just need to apply the existing models. As a start, they suggest you can look at small datasets using GUI-based software like Weka. Python also has extensive libraries like Keras (keras.io). And there are open source Machine Learning libraries like Tensorflow and Theano.

Figure 41: Typical tasks and frequently used Machine Learning methods

Question	Data Analysis Technique
Given set of inputs, predict asset price direction	Support Vector Classifier, Logistic Regression, Lasso Regression, etc.
How will a sharp move in one asset affect other assets?	Impulse Response Function, Granger Causality
Is an asset diverging from other related assets?	One-vs-rest classification
Which assets move together?	Affinity Propagation, Manifold Embedding
What factors are driving asset price?	Principal Component Analysis, Independent
Is the asset move excessive, and will it revert?	Component Analysis
What is the current market regime?	Soft-max classification, Hidden Markov Model
What is the probability of an event?	Decision Tree, Random Forest
What are the most common signs of market stress?	K-means clustering
Find signals in noisy data	Low-pass filters, SVM
Predict volatility based on a large number of input variables	Restricted Boltzmann Machine, SVM
What is the sentiment of an article / text?	Bag of words
What is the topic of an article/text?	Term/InverseDocument Frequency
Counting objects in an image (satellite, drone, etc)	Convolutional Neural Nets
What should be optimal execution speed?	Reinforcement Learning using Partially Observed Markov Decision Process

Source: J.P. Morgan Macro Q&A

Know what you are doing



Solving the Problem

How do you allow for impossible events?

And what is, actually, impossible?

—*‘that never really happens’*

—*‘It makes no sense, it cannot happen’*

—*‘That is just mathematically impossible’*

We will not mess with the
mathematically impossible...

'That never really happens'

Once upon a time in 2007...

*No, equities
always grow at
positive rates in
the long term.*

*Our profit testing
assumes equity growth
is always positive,
should we also test
when it is not?*



'That never really happens'

Later, the following year...

Page last updated at 21:32 GMT, Wednesday, 31 December 2008

 E-mail this to a friend

 Printable version

FTSE 100 index has its worst year

The financial brutality of 2008 has been confirmed after the FTSE 100 index recorded its biggest annual decline since its inception in 1984.

Britain's main share index ended 2008 trading down 31.3% compared with a year earlier.

With trading closing at 1230 GMT, the FTSE finished 2008 at 4,434 points. A year ago it closed at 6,457.

In New York, the Dow Jones closed slightly up, but was down almost 34% for 2008, its worst year since 1931.

The FTSE indexes, including the FTSE 250, are the benchmark for investors, including institutional funds.

BBC: <http://news.bbc.co.uk/2/hi/business/7805008.stm>



The FTSE 100 has had to endure a turbulent time in 2008



'It makes no sense, it cannot happen'

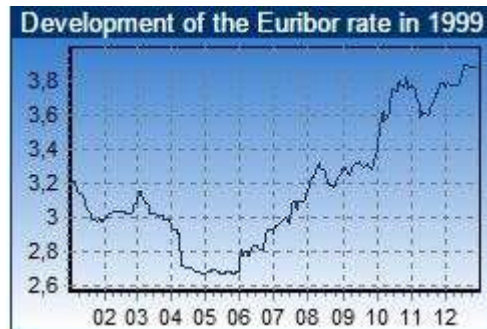
Once upon a time in 2010...

*This 20 year
simulation considers
only positive interest
rates, should we think
of negative too?*

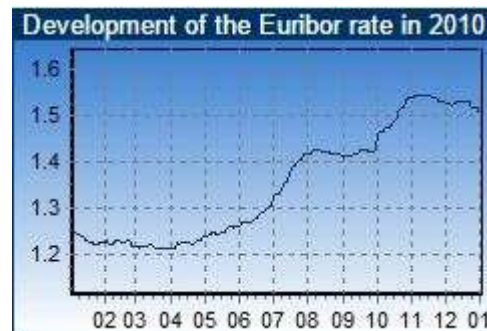
*Negative interest
rates? That is just
stupid. They are
already at their
lowest ever.*



'It makes no sense, it cannot happen'



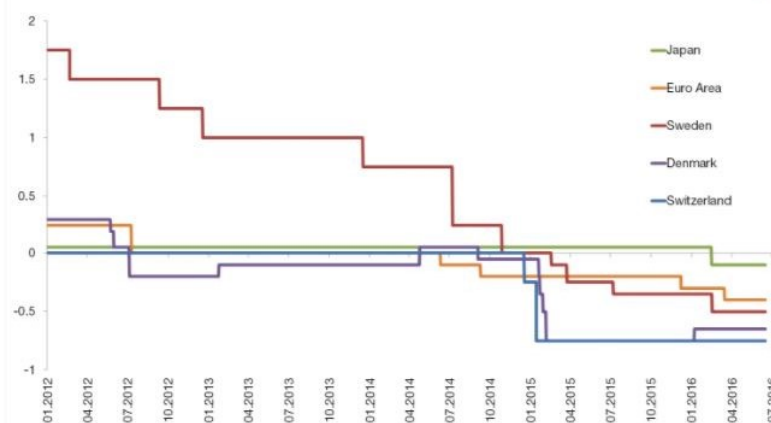
You see? Over 2.6% in 1999 and at least 1.2% now in 2010. It cannot get lower than that. Negative rates make no economic sense.



'It makes no sense, it cannot happen'

But in 2016...

The fall of key market interest rates below zero



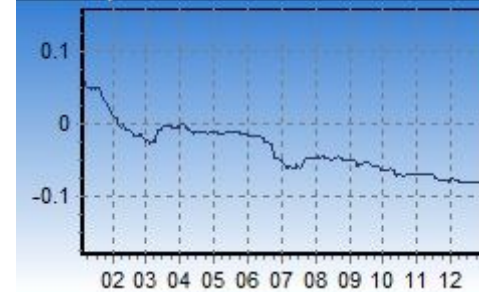
Source: National Central Banks, provided by Haver Analytics

A short history

Throughout history, it was widely believed that central banks could not move short-term interest rates below zero. After all, why would anyone pay to deposit money in a bank or pay to lend someone money, when they could just keep their cash at home for free? Cash always has a zero interest rate.

It was widely believed that if interest rates did dip below zero, even if by a very small amount, everyone with savings would run to the bank to change them for ready money. The zero interest rate on cash was seen as the lowest point an interest rate could dip to, the point at which central banks would be out of ammunition.

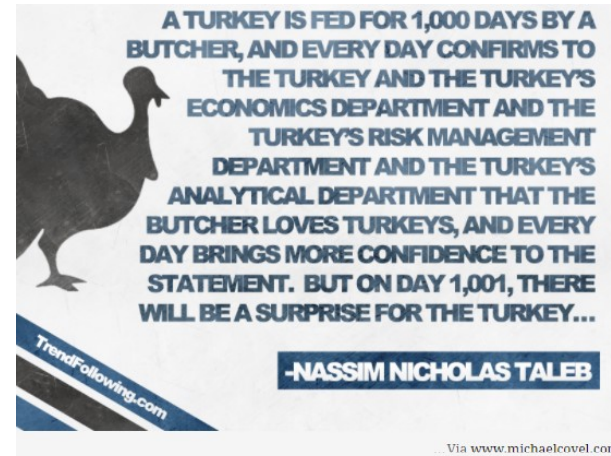
Development of the Euribor rate in 2016



World Economic Forum: <https://www.weforum.org/agenda/2016/11/negative-interest-rates-absolutely-everything-you-need-to-know/>

Solving the Problem

- Be sceptical and even question your own wisdom.
- Consider scenarios beyond what seems possible. What is not possible under current circumstances could easily change.
- Prepare to be laughed at. They will think that you are a Muppet!



1696 – Willem de Vlamingh is utterly convinced that all swans are white. Everybody knows it as there have not been any reports of non-white swans in 4,000 years.

1697 – Willem de Vlamingh tells the world that swans can actually be black. He has seen them in Australia.

2017 – Would your model consider the possibility that a swan may be pink? What if I told you I have seen one?

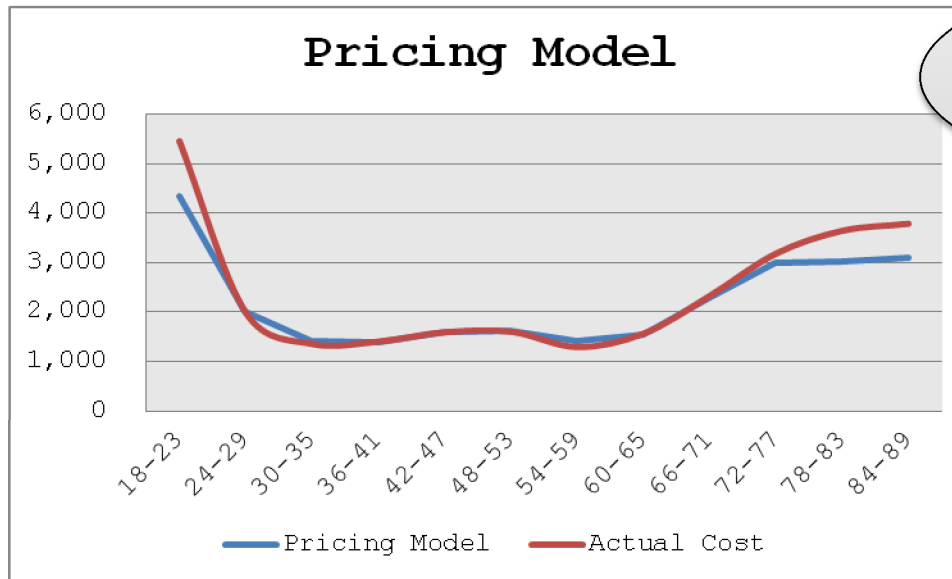
Deliver the Solution

Look at this, we are losing customers. Do you think we could price our book more competitively?



I will see what I can do before I go on holidays.

Deliver the Solution



Age Band	Pricing Model	Actual Cost	Error	Error Sq
18-23	4,354	5,457	1,103	1,216,609
24-29	2,015	1,997	-18	324
30-35	1,427	1,365	-62	3,844
36-41	1,399	1,413	14	196
42-47	1,605	1,599	-6	36
48-53	1,622	1,613	-9	81
54-59	1,412	1,303	-109	11,881
60-65	1,534	1,571	37	1,369
66-71	2,301	2,333	32	1,024
72-77	2,998	3,187	189	35,721
78-83	3,023	3,648	625	390,625
84-89	3,107	3,789	682	465,124
			2,478	2,126,834

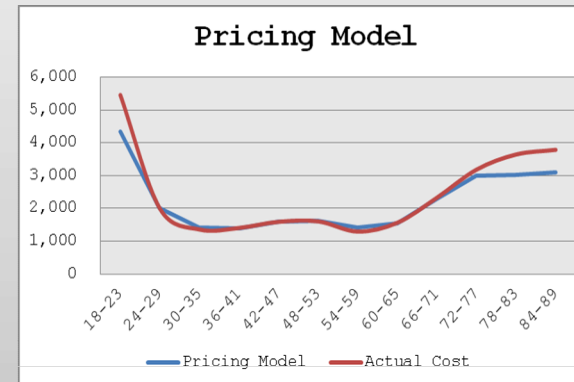
That looks really good, I will take it to the Board while you are away.



Sure, just remember it does not work well with the young and the elderly but works well in average.

Deliver the Solution

The Board considered it while you were on holidays and decided to focus on young and elderly customers, as that is where we can be more competitive.



Deliver the Solution

- Could we have communicated the model's shortcomings more clearly?
- Should we have insisted on presenting the model ourselves?
- Should we have capped the model so that we do not produce results where the error is too high, even if it works well on average?
- Why can they never listen to the detail? That's where the devil is!

And the audience cheerfully participates once again

“All you need in life is ignorance and confidence, and success is sure.”

Mark Twain

And Be the Hero

