

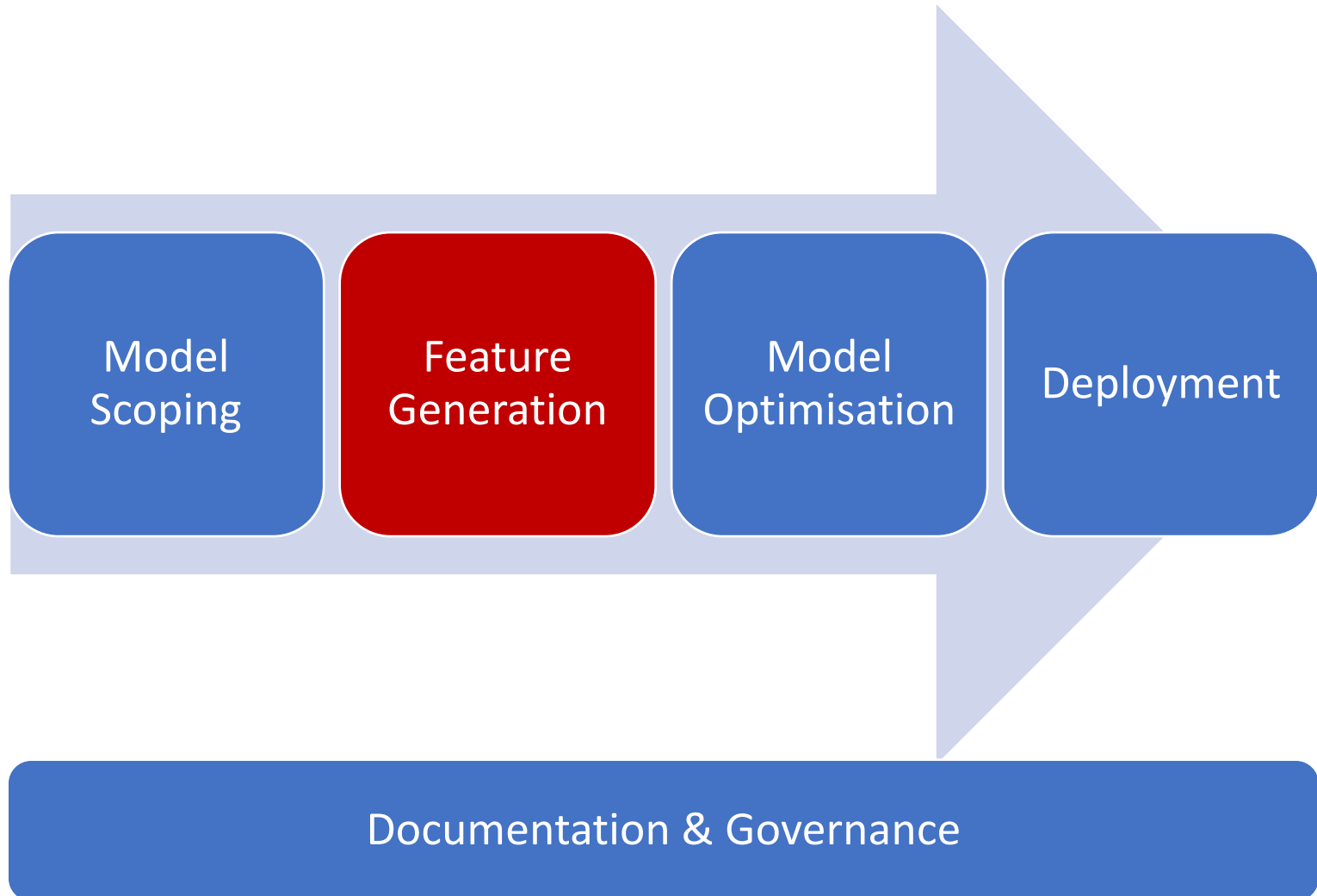


A Comparison of Feature Generation Techniques for Credit Risk Modelling

Dan Kellett

Director of Data Science

Feature Generation is a core part of modelling but is often overlooked for more 'glamorous' parts of the project



A move to tradeline bureau data dramatically expanded the potential for new feature creation

ACCT_ID	Ave balance	Worst status	# Searches
12345	500	0	0
38410	120	1	3
86500	1750	8	1

1 row per customer



ACCT_ID	Account	Balance	Status
12345	1	250	0
12345	2	100	1
12345	3	0	0

1 row per customer, per bureau record

The expert driven approach uses innovation techniques to access an organisation's bank of knowledge



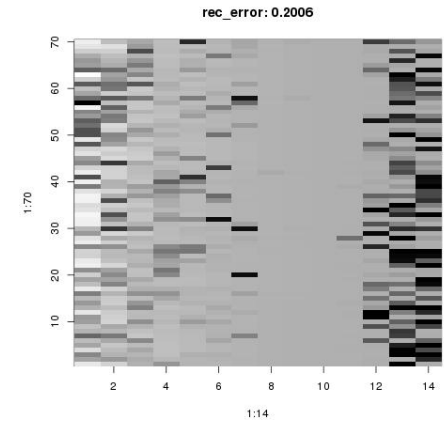
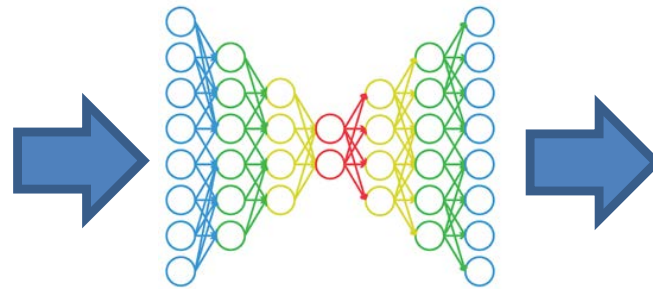
The brute force approach is about rapid progression through a feature extraction funnel

- Mechanistically created **1.4 million** new variables from bureau data for risk prediction
- Just **15 hours** processing time reduced this to **12** powerful new splitters
- The app-stage risk model contains **4** of these covariates as some of the top variables

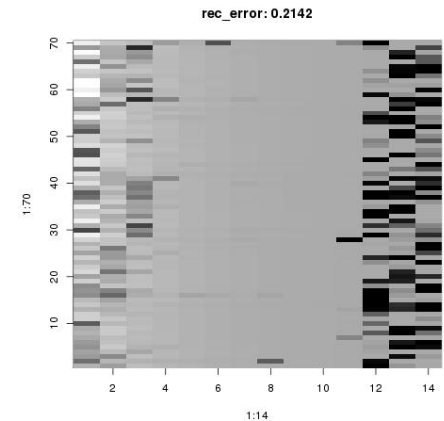
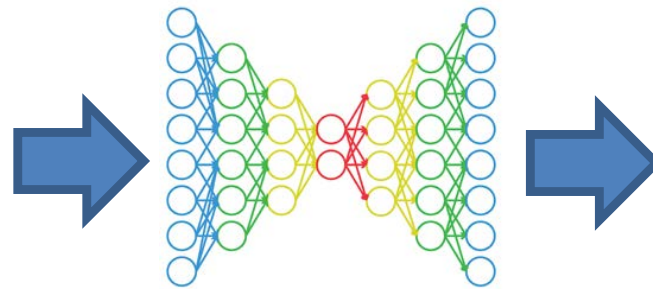


The AI approach uses Deep Learning to build pictures of 'typical' risk profiles for direct comparison

Bureau-gram
of low risk
individual



Bureau-gram
of high risk
individual





Dan Kellett

@kellettboy1 | @CapitalOneGeek