Risk and Artificial Intelligence: Credit Scoring in Practice

Dr. Paul Larsen

April 15, 2022

Operational Risk in Credit Scoring

Model Risk and Process Risk

The AI Act

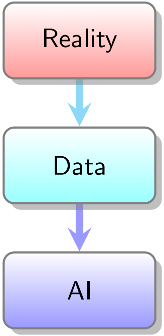
Manage model risk: Don't let the data-fed models speak nonsense

The Supervisory Requirements for IT in German Asset Managers

Manage process risk: Automate everything that makes sense

Manage Model Risk

Don't outsource thinking to algorithms



Manage Model Risk

Example of credit scoring for loans

Artificial credit scoring data sample:

gender	occupation	activity	default
1	0	0	1
0	0	0	0
0	0	1	1
0	0	0	0
0	1	0	0

gender: 0 for female, 1 for male [1]
occupation: 0 for education, 1 for health
activity: 0 for low account activity, 1 for high account activity
default: 0 for no-default, 1 for default

[1] Some current and future credit data will likely have a non-binary gender category.

Manage Model Risk

Simpson's paradox again

Correlations with default:

pair	pearson_r	p_value
('gender', 'default')	-0.097396	6.715668e-10
('occupation', 'default')	-0.221915	8.141972e-46
('activity', 'default')	0.953242	0.000000e+00

Selected sub-population default rates:

population	female	male
total	0.395094	0.302303
education	0.449205	0.461679
health	0.228145	0.245443

Manage Model Risk

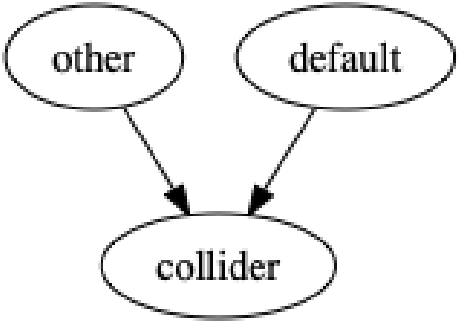
Spurious correlations aren't the only ones to fear

- Colliders from target to a feature(s)
- Data leakage

Manage Model Risk

Collider risk: causality knows direction, correlation doesn't

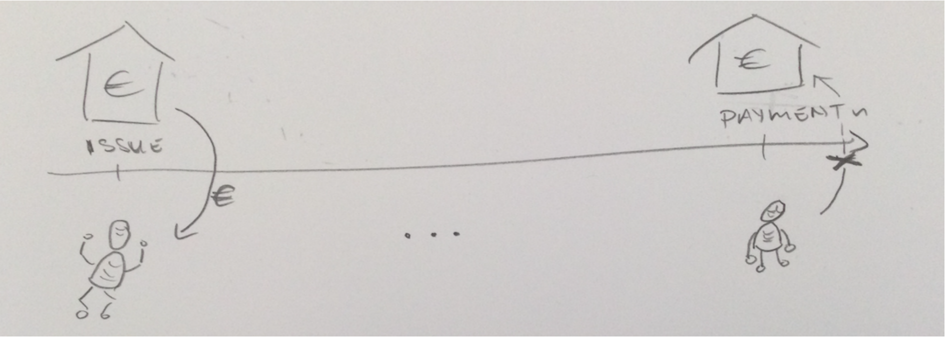
What if we had causal graph with features $\mathrm{collider}, \mathrm{other}$ and target $\mathrm{default}$?



Manage Model Risk

Data leakage risk: know your process, know your data

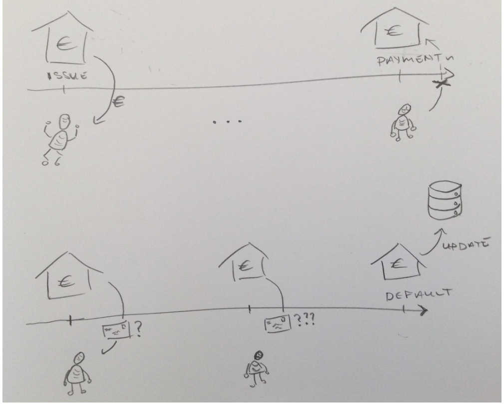
How I understood credit default pre-practical-experience



Manage Model Risk

Data leakage risk II: know your process, know your data

How credit default really[2] happens



[2] more accurately, a less violent projection of reality

Manage Process Risk

Working with human nature in software development

Source:
https://commons.wikimedia.org/wiki/File:Codex_Aemilianensis.jpg

- Agile: agilemanifesto.org
- Test Driven Development: Kent Beck, *Test Driven Development: By Example*
- DevOps and MLOps: Google's *Site Reliability Engineering*

Manage Process Risk

Isn't AI software different?

Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.

Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison
{ebner,vchaudhary,mwyoung,jfcrespo,dennison}@google.com
Google, Inc.

Abstract

Machine learning offers a fantastically powerful toolkit for building useful complex prediction systems quickly. This paper argues it is dangerous to think of these quick wins as coming for free. Using the software engineering framework of *technical debt*, we find it is common to incur massive ongoing maintenance costs in real-world ML systems. We explore several ML-specific risk factors to account for in system design. These include boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, configuration issues, changes in the external world, and a variety of system-level anti-patterns.

Source: Neurips 2015, Hidden TechDebt in ML

***"Everyone wants to do the model work, not the data work"*: Data Cascades in High-Stakes AI**

Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, Lora Aroyo
[nithyasamba,kapania,hhighfill,dakrong,pkp,loraa]@google.com
Google Research
Mountain View, CA

ABSTRACT

AI models are increasingly applied in high-stakes domains like health and conservation. Data quality carries an elevated significance in high-stakes AI due to its heightened downstream impact, impacting predictions like cancer detection, wildlife poaching, and loan allocations. Paradoxically, data is the most under-valued and de-glamorised aspect of AI. In this paper, we report on data practices in high-stakes AI, from interviews with 53 AI practitioners in India,

lionized work of building novel models and algorithms [46, 125]. Intuitively, AI developers understand that data quality matters, often spending inordinate amounts of time on data tasks [60]. In practice, most organisations fail to create or meet any data quality standards [87], from under-valuing data work vis-a-vis model development.

Under-valuing of data work is common to all of AI development [125][1]. We pay particular attention to undervaluing of data in *high-stakes domains*[2] that have safety impacts on living beings,

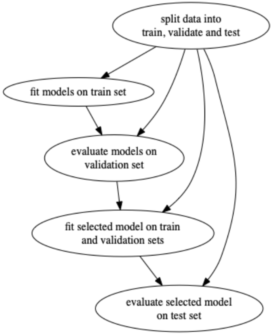
Source: Data Cascades in High-Stakes AI

Manage Process Risk

Matching low-resolution buzzwords with high-resolution practice

Continuous Integration, Continuous Delivery for Machine Learning

(Semi-)automated pipelines for development



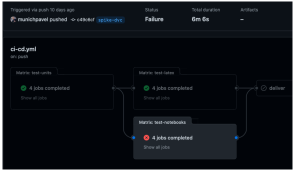
derived from risk_leaning/model_selection/run-pipeline.py

Manage Process Risk

Matching low-resolution buzzwords with high-resolution practice

Continuous Integration, Continuous Delivery for Machine Learning, II

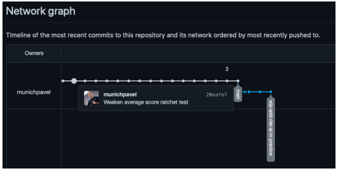
(Semi-)automated pipelines for production



Source: this repo's CI CD run #37; see CI CD #75 for a successful pipeline run, and the other CI CD pipeline runs for many other near-misses.

Version control

Git (or cousin) for collaboration sanity and audit-trail

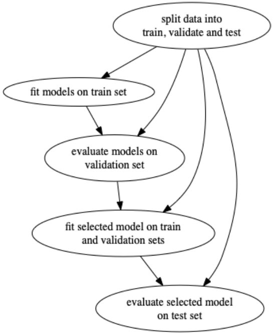


Source: This repo's GitHub commit graph

Wrapping up

How did the credit scoring model perform?

Our credit scoring AI pipeline:



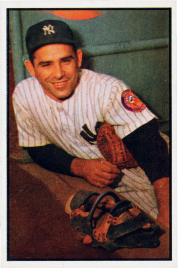
On the validation set:

- Logistic regression scores
- Decision tree scores
- Random forest scores

On the test set: Chosen model scores

Wrapping up, II

Don't worry: AI won't replace you soon



In theory, there is no difference between theory and practice. In practice, there is.
Yogi Berra

→ Risk is to managed, not eliminated

→ Manage operational risk in AI by think deeply in practice

…about the process and its data
…about the algorithms you use
…about potential impacts on society

→ Manage operational risk in AI by automating everything sensible