

# The AI Playbook

## Mastering the Rare Art of Machine Learning Deployment

By Eric Siegel

### Notes for CHAPTER 4

These notes include references, plus resources for further learning. For all the chapters' notes as well as information about the book in general, access the book's website at [www.bizML.com](http://www.bizML.com).

---

A technical note on training data being a flat table ("BOFF"): Some ML software can directly access data from across multiple tables rather than requiring a pre-formed table, but it must be provided the conceptual definition or "view" of a flat, two-dimensional training dataset just the same.

A nice thread on the under-prioritization of data prep:

<https://www.linkedin.com/feed/update/urn%3Ali%3Aactivity%3A7024769288062341120/>

Card Fraud Losses Dip to \$28.58 Billion (2021)

<https://nilsonreport.com/mention/1515/1link/>

Nilson Report, Issue 1209, December 2021

[https://nilsonreport.com/upload/content\\_promo/NilsonReport\\_Issue1209.pdf](https://nilsonreport.com/upload/content_promo/NilsonReport_Issue1209.pdf)

Card industry faces \$400B in fraud losses over next decade, Nilson says

<https://www.paymentsdive.com/news/card-industry-faces-400b-in-fraud-losses-over-next-decade-nilson-says/611521/>

Card fraud overview

<https://youtu.be/c-DxF1XVATw>

Global Fraud Trends... Welcome to the 'Scampocalypse':

<https://www.fico.com/blogs/global-fraud-trends-whats-happening-and-whats-next>

FICO Machine Learning Algorithms Improve Card-Not-Present Fraud Detection by 30%

<https://www.prnewswire.com/news-releases/fico-machine-learning-algorithms-improve-card-not-present-fraud-detection-by-30-300529629.html>

Elo Uses FICO Falcon Fraud Manager to Reduce Fraud by 30% Across Its Payment Network  
<https://www.businesswire.com/news/home/20220531005048/en/Elo-Uses-FICO-Falcon-Fraud-Manager-to-Reduce-Fraud-by-30-Across-Its-Payment-Network>

How FICO Has Made A Fortune Off Your Credit Score  
<https://www.forbes.com/sites/jonmarkman/2019/08/27/how-fico-has-made-a-fortune-off-your-credit-score/>

FICO Takes Its Analytics Expertise Far Beyond Credit Scores  
<https://www.forbes.com/sites/oracle/2019/05/01/fico-takes-its-analytics-expertise-far-beyond-credit-scores/>

The Data Economy Podcast: Machine Learning Models for Real-Time Financial Services with FICO  
<https://www.youtube.com/watch?v=oshYnZYnvkl>

FICO Falcon Fraud Manager for Retail Banking (2010)  
<https://docplayer.net/17002316-Fico-falcon-fraud-manager-for-retail-banking.html>

Small business loan risk informed by consumer-level data:  
From "Managing Data Mining: Advice from Experts" by Stephan Kudyba, in the chapter "Credit Scoring and Risk Management for Small Business Lending", by Vernon Gerety.

The "data leak" example for an insurance company is from the book "Calling Bullshit: The Art of Skepticism in a Data-Driven World," by Carl Bergstrom and Jevin West.

It is a mistake to accept leaks from the future  
<https://www.elderresearch.com/blog/avoid-leaks-from-future>

2021 State of Data Science survey by Anaconda:  
<https://know.anaconda.com/rs/387-XNW-688/images/Anaconda-2021-SODS-Report-Final.pdf>

The Data Economy Podcast: Machine Learning Models for Real-Time Financial Services with FICO  
<https://www.youtube.com/watch?v=oshYnZYnvkl>

FICO Falcon Fraud Manager  
<https://community.fico.com/s/fraud?tabset-26fe7=2>

Wise Practitioner – Predictive Analytics Interview Series: Scott Zoldi at FICO, in The Machine Learning Times  
<https://www.predictiveanalyticsworld.com/machinelearningtimes/wise-practitioner-predictive-analytics-interview-series-scott-zoldi-fico08192016/8001/>

FICO Takes Its Analytics Expertise Far Beyond Credit Scores

<https://www.forbes.com/sites/oracle/2019/05/01/fico-takes-its-analytics-expertise-far-beyond-credit-scores/>

Quotation from Major General C. H. White:

<https://quoteinvestigator.com/2017/02/23/judgment/>

Data-Centric AI Competition

<https://https-deeplearning-ai.github.io/data-centric-comp/>

<https://www.deeplearning.ai/blog/>

<https://datacentricai.org>

Virtual Sweatshops Paint A Bleak Picture Of The Future Of Work

<https://web.archive.org/web/20220117232043/https://www.vocativ.com/410794/are-virtual-sweatshops-the-future-of-work/index.html>

The Human Labor Behind Artificial Intelligence

<https://www.marketplace.org/2021/05/04/the-human-labor-behind-artificial-intelligence/>

Resolving the Human Subjects Status of Machine Learning's Crowdworkers

<https://arxiv.org/abs/2206.04039>

The Exploited Labor Behind Artificial Intelligence

<https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/>

The Human Bots Who Power Parasite Platforms

<https://points.datasociety.net/the-human-bots-who-power-parasite-platforms-c165fd94fe35>

Two more examples of merging in new columns of data (input variables):

- For EduPay, I merged in Census information about the user's neighborhood, based on zip-code level data. This included the population (reflective of whether it's an urban or suburban area), geographical categories, and even the elevation.

- A credit scoring project for small businesses – where the model predicts whether each business would default on their loan – discovered the value of introducing information about the humans behind the corporation. By introducing inputs that summarized the business owner's prior behavior – based on consumer-level data about that individual – the resulting model improved. In fact, a model based on consumer data alone was better than one based only on commercial-level data; a model based on both data sources worked best of all. This improved the model's lift at 30% from 1.8 up to 2.2. When managing business loans, incorporating consumer data is a break from norms, but one that can help.

