The AI Playbook

Mastering the Rare Art of Machine Learning Deployment

By Eric Siegel

Notes for CHAPTER 3

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.

THE ACCURACY FALLACY:

The following article on the accuracy fallacy by the author covers similar material to Chapter 3 and then continues further with numerous additional examples:
https://www.predictiveanalyticsworld.com/machinelearningtimes/accuracy-fallacy-the-medias-coverage-of-ai-is-bogus/10652/

Two-part video on the accuracy fallacy by the author:

Accuracy fallacy: the media's bogus coverage of ML – from "Machine Learning Leadership and Practice"
https://www.youtube.com/watch?v=81Vv0J2Vw-Y

Accuracy fallacy #2: predict psychosis, criminality, bestsellers – from "ML Leadership and Practice"
https://www.youtube.com/watch?v=ui3VkecTX3Y

Podcast episode on the accuracy fallacy by the author:
https://doctordatashow.com/e/the-accuracy-fallacy-and-the-media-s-bogus-coverage-of-ai/

- - - - - - - - - - - -

Thank you to Scientific American, which originally published my article on the accuracy fallacy, an earlier version of Chapter 3's opening:
https://blogs.scientificamerican.com/observations/the-medias-coverage-of-ai-is-bogus/

Thank you to Big Think, which originally published a version of Chapter 3's section on how probabilities such as Nate Silver's are often misunderstood:

https://bigthink.com/the-present/election-prediction/

A caveat regarding the accuracy fallacy: Accuracy as a metric is more often potentially helpful if you move beyond the binary (yes/no) prediction goals on which this book mostly focuses to multiple-category problems, such as predicting which university a student will attend.

Another issue with the Stanford "gaydar" report is that the researchers assert that their results support a hormonal theory of sexuality rather than solely reflecting cultural trends in self-presentation. This problematic claim has been duly debunked:
https://www.callingbullshit.org/case_studies/case_study_ml_sexual_orientation.html

- - - - - - - - - - - -

Newsweek: “AI Can Tell If You’re Gay: Artificial Intelligence Predicts Sexuality from One Photo with Startling Accuracy”
https://www.newsweek.com/ai-can-tell-if-youre-gay-artificial-intelligence-predicts-sexuality-one-photo-661643

The Spectator: “Linguistic Analysis Can Accurately Predict Psychosis”
https://web.archive.org/web/20210126223118/https://life.spectator.co.uk/articles/linguistic-analysis-can-accurately-predict-psychosis/

The Daily Mail: “AI-Powered Scans Can Identify People at Risk of a Fatal Heart Attack Almost a Decade in Advance… with 90% Accuracy”
https://www.dailymail.co.uk/health/article-7422565/AI-powered-scans-spot-heart-attack-decade-strikes.html

The Next Web: “This Scary AI Has Learned How to Pick Out Criminals by Their Faces”
https://thenextweb.com/news/artificial-intelligence-criminals-face

Stanford University’s infamous “gaydar” study:
"Deep neural networks are more accurate than humans at detecting sexual orientation from facial images," by Michal Kosinski and Yilun Wang, Journal of Personality and Social Psychology, 2018, Vol. 114, No. 2, 246–257
https://psyarxiv.com/hv28a/

AI Can Tell If You're Gay: Artificial Intelligence Predicts Sexuality From One Photo with Startling Accuracy
https://www.newsweek.com/ai-can-tell-if-youre-gay-artificial-intelligence-predicts-sexuality-one-photo-661643

The Economist's coverage and its related front cover illustration
https://drive.google.com/file/d/1eDvkTTgbaf4hHuhMXFC1Oj6kaDP9ZtD8/view

https://www.economist.com/science-and-technology/2017/09/09/advances-in-ai-are-used-to-spot-signs-of-sexuality

The AUC is mathematically equal to the result you get running the pairing test
https://stats.stackexchange.com/questions/180638/how-to-derive-the-probabilistic-interpretation-of-the-auc

“AUC: A Fatally Flawed Model Metric," by John Elder
https://www.elderresearch.com/blog/auc-a-fatally-flawed-model-metric

Machine learning algorithms can help psychologists predict, with 90% accuracy, the onset of psychosis by analyzing a patient’s conversations.
https://www.theregister.com/2019/06/18/ai_algorithms_psychosis/

Nature publication thereof
https://www.nature.com/articles/s41537-019-0077-9

The general population exhibits a 3% rate of psychotic disorders
https://pubmed.ncbi.nlm.nih.gov/17199051/

A researcher claims to “accurately predict a patient’s first psychiatric hospitalization and diagnosis, more than a year before it happened” based on Facebook activities
https://www.psychologytoday.com/us/blog/the-health-our-youth/202012/inferring-psychiatric-illness-based-digital-activity-crosses

The belief held by some psychiatrists that AI will replace their job
https://www.businesswire.com/news/home/20190730005743/en/Global-Survey-Psychiatrists-Impact-Artificial-Intelligence

The Global Times (China): “Professor Claims AI Can Spot Criminals by Looking at Photos 90% of the Time”
https://web.archive.org/web/20201111125246/https://www.globaltimes.cn/content/1026928.shtml

Repeat of that claim:
https://www.technologyreview.com/2016/11/22/107128/neural-network-learns-to-identify-criminals-by-their-faces/
https://www.telegraph.co.uk/technology/2016/11/24/minority-report-style-ai-learns-predict-people-criminals-facial/

Researchers' original publication
https://arxiv.org/pdf/1611.04135v2.pdf

“Google AI Predicts Hospital Inpatient Death Risks with 95% Accuracy”

https://www.fiercebiotech.com/medtech/google-ai-predicts-hospital-inpatient-death-risks-95-accuracy

Google researchers published this result in Nature
https://www.nature.com/articles/s41746-018-0029-1

A model “that predicted suicide risk, using electronic health records, with 84 to 92% accuracy within one week of a suicide event.”
https://phys.org/news/2018-03-ai-suicides.html

Vanderbilt University research publication
https://journals.sagepub.com/doi/abs/10.1177/2167702617691560

How AI is helping to predict and prevent suicides
https://phys.org/news/2018-03-ai-suicides.html

Predicting Risk of Suicide Attempts Over Time Through Machine Learning
https://journals.sagepub.com/doi/abs/10.1177/2167702617691560

IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs
https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html

80% accuracy claim from “The Bestseller Code: Anatomy of the Blockbuster Novel”
https://www.theguardian.com/money/2017/sep/23/write-bestselling-novel-algorithm-earning-money
https://www.wsj.com/articles/an-algorithm-to-predict-a-bestseller-1472659425
https://www.independent.co.uk/arts-entertainment/books/news/the-bestseller-code-the-words-that-make-a-successful-book-a7117211.html

A manuscript predicted by the model as a “future bestseller” actually has less than a 2% probability
https://medium.com/@arhomberg/is-it-possible-to-predict-the-next-nyt-bestseller-34f7b28c5181

FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals
https://arxiv.org/pdf/1901.02212.pdf

Artificial intelligence could be used to accurately predict tsunamis
https://www.cardiff.ac.uk/news/view/2587345-artificial-intelligence-ai-could-be-used-to-accurately-predict-tsunamis

AI-powered scans can identify people at risk of a fatal heart attack almost a DECADE in advance

https://www.dailymail.co.uk/health/article-7422565/AI-powered-scans-spot-heart-attack-decade-strikes.html

- - - - - - - - - - - -

Despite its pervasiveness, plenty of practitioners instead prefer integrity to hype. Analytics entrepreneur Scott Carl criticizes the accuracy fallacy: "This exaggeration is bad for the practitioner and bad for the industry and very bad for the client."

How Nate Silver's U.S. Presidential Forecasts Were Unfairly Criticized:
https://bigthink.com/the-present/election-prediction/
https://fivethirtyeight.com/features/election-update-why-our-model-is-more-bullish-than-others-on-trump/
https://news.harvard.edu/gazette/story/2017/03/nate-silver-says-conventional-wisdom-not-data-killed-2016-election-forecasts/

Corresponding with the sidebar "The Profit of Response Modeling," here is a spreadsheet detailing the calculations, which you may copy and toy with at will. NOTE - PLEASE DO NOT REQUEST PERMISSION TO EDIT; INSTEAD, MAKE A COPY THAT YOU CAN EDIT.
https://docs.google.com/spreadsheets/d/1sLp2sGxTZKH0FW4x-ViukfZ-RsCYDrD9B-ReuMZus8M/edit?usp=sharing

Quote from Wafiq Syed:
The Most Important Metric for Machine Learning Models
https://towardsdatascience.com/the-most-important-metric-for-machine-learning-models-2c6a4c4b18ad

Quote from Mark Tenenholtz:
https://mobile.twitter.com/marktenenholtz/status/1506964069470810119

Quote from Katie Malone:
When Translation Problems Arise Between Data Scientists and Business Stakeholders, Revisit Your Metrics
https://hdsr.mitpress.mit.edu/pub/bfeyfx22/release/2

Sometimes the triage achieved by ML is literally medical triage:
http://open.spotify.com/episode/6J03LJxFcsIAvwROaYSWkI?si=zS08pWXiQZe78VMCegDKbA

Citizens Bank checking fraud example
Jay Zhou, “Building In-Database Predictive Scoring Model: Check Fraud Detection Case Study,” Predictive Analytics World Washington, DC, Conference, October 20, 2009, Washington, DC.
https://www.predictiveanalyticsworld.com/dc/2009/agenda.php#day1%E2%80%9319

Six ways to lower costs with predictive analytics

https://www.the-digital-insurer.com/wp-content/uploads/2013/12/85-Six-Ways-to-Lower-Costs-with-Predictive-Analytics.pdf
https://www.predictiveanalyticsworld.com/lower-costs-with-predictive-analytics.php

Counterpoint: "AI success comes through growth, not labor savings"
https://www.forbes.com/sites/joemckendrick/2018/10/25/artificial-intelligence-success-come-through-growth-not-labor-savings/

Spreadsheet with the fraud example's calculations so that you can try out changes to the scenario at will. **NOTE - PLEASE DO NOT REQUEST PERMISSION TO EDIT; INSTEAD MAKE A COPY THAT YOU CAN EDIT.**
https://docs.google.com/spreadsheets/d/12dKRw4TYmqwbAkp3VRTwt5gUe1rnLweg0EHfRBQHcIc/edit?usp=sharingg

"9 Laws of Data Mining" by Tom Khabaza, 2022 (self-published book). Page 87.

The Most Important Metric for Machine Learning Models
"The true measure of a model's success is how it impacts the business."
https://towardsdatascience.com/the-most-important-metric-for-machine-learning-models-2c6a4c4b18ad

The false positive rates of at-home Covid tests:
https://www.bmj.com/content/373/bmj.n1411/rr
https://www.self.com/story/the-view-how-common-false-positive-covid-19-tests
https://www.fda.gov/medical-devices/letters-health-care-providers/potential-false-positive-results-antigen-tests-rapid-detection-sars-cov-2-letter-clinical-laboratory

One popular test misses around 15 out of 100 infections and gives a FP result in about 1 in 100 people who aren't infected.
https://www.nbcchicago.com/news/coronavirus/how-accurate-are-at-home-covid-tests-what-to-know-ahead-of-thanksgiving/2693121/

An in-depth dive into machine learning ethics by the author of this book, Eric Siegel -- articles and videos:
http://www.civilrightsdata.com

Six ethical quandaries of predictive policing
https://www.youtube.com/watch?v=y9oI8Uy3Xp8

"Even the best decision doesn't yield the best outcome every time. There's always an element of luck that you can't control..."
-Annie Duke, poker champion

"Why is the mean squared error on the predicted number of stars an appropriate metric for our recommendation problem? Is it meaningful? Is there a better metric? Hopefully, the analyst has thought this through carefully. It is surprising how often one finds that an analyst has not and is simply reporting some measure he learned about in a class in school."
- Foster Provost and Tom Fawcett, "Data Science for Business"

“Your model is worthless if you can't communicate its value.”
–Mark Tenenholtz, data science entrepreneur and Kaggle Master

ALGORITHMIC BIAS:

The foundational, often-cited article that started the conversation about algorithmic bias (aka machine bias):
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Can you make AI fairer than a judge? Play our courtroom algorithm game
https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/

Researchers at Google present another interactive visualization, this time regarding flagging loan applicants for defaults (rather than criminal defendants) and depicting, comparing, and contrasting other proposed fairness standards:
Attacking discrimination with smarter machine learning
https://research.google.com/bigpicture/attacking-discrimination-in-ml/