

Healthcare Machine Learning

with

healthcare.ai



Josh O'Rourke

About Me

- Senior Software Developer at Xtend Healthcare
- Very recent transplant to the Nashville area
- Director of Solution Architecture at King's Daughters health system
- Led the implementation of Health Catalyst BI Platform
- Learned about healthcare.ai and became a contributor

I ❤️ Healthcare



AI & Healthcare



THE HYPE

Healthcare and High Tech



IBM

Diagnosis, Genomics,
and Oncology



Google

Calico, DeepMind
Health, Verily



Amazon

EMR, Telemedicine,
Pharmaceuticals



Artificial Intelligence

IBM's Watson AI Recommends Same Treatment as Doctors in 99% of Cancer Cases

David Ramos/Getty Images

IN BRIEF

- Watson recommended treatment plans that matched suggestions from oncologists in 99 percent of the cases it analyzed and offered options doctors missed in 30 percent of them.
- AI could be revolutionary for healthcare as it can process many more research papers and case files than any human doctor could manage.

SHARE



WRITTEN BY

Dom Galeon

@domgaleon



Kristin Houser
Website



October 28, 2016

How 3D Printing and IBM Watson Could Replace Doctors



FORTUNE

BRAINSTORM

HEALTH

Artificial intelligence might make fewer mistakes than your doctor, Athenahealth's CEO said.

By [Jen Wieczner](#) Nov 2nd, 2016 12:51 AM ET

Can a supercomputer with artificial intelligence take better care of you than your doctor can?

Stories From



**Jim Acosta Calls
Trump's Behavior
Exactly What It Was**

Ideal

**Leaked Memo
Reveals White
House Strategy For
Puerto Rico**

A Reality Check for IBM's AI Ambitions

IBM, number 39 on our list of the 50 Smartest Companies, overhyped its Watson machine-learning system, but the company still could have the best access to the kind of data needed to make medicine much smarter.

by David H. Freedman June 27, 2017



LEONARD GRECO



“ We do ourselves a disservice when we refer to AI as something that we just take a ton of data, push a button and cure cancer. We need to be more realistic about what these tools can and can’t do. ”

Leonard D’Avolio



“ Many people today like to imagine a world where AI can replace doctors. But the problem with the U.S. healthcare system isn't our providers — people fly from around the world to get treated by them. Instead, it's the ability to reliably provide an operational environment that allows these world-class clinicians to do their best work. This may seem boring and unsexy compared to the “robo-surgeon” but it's where we have the biggest opportunities to improve the patient experience and reduce cost of delivering care.

Mudit Garg, CEO of Qvetus



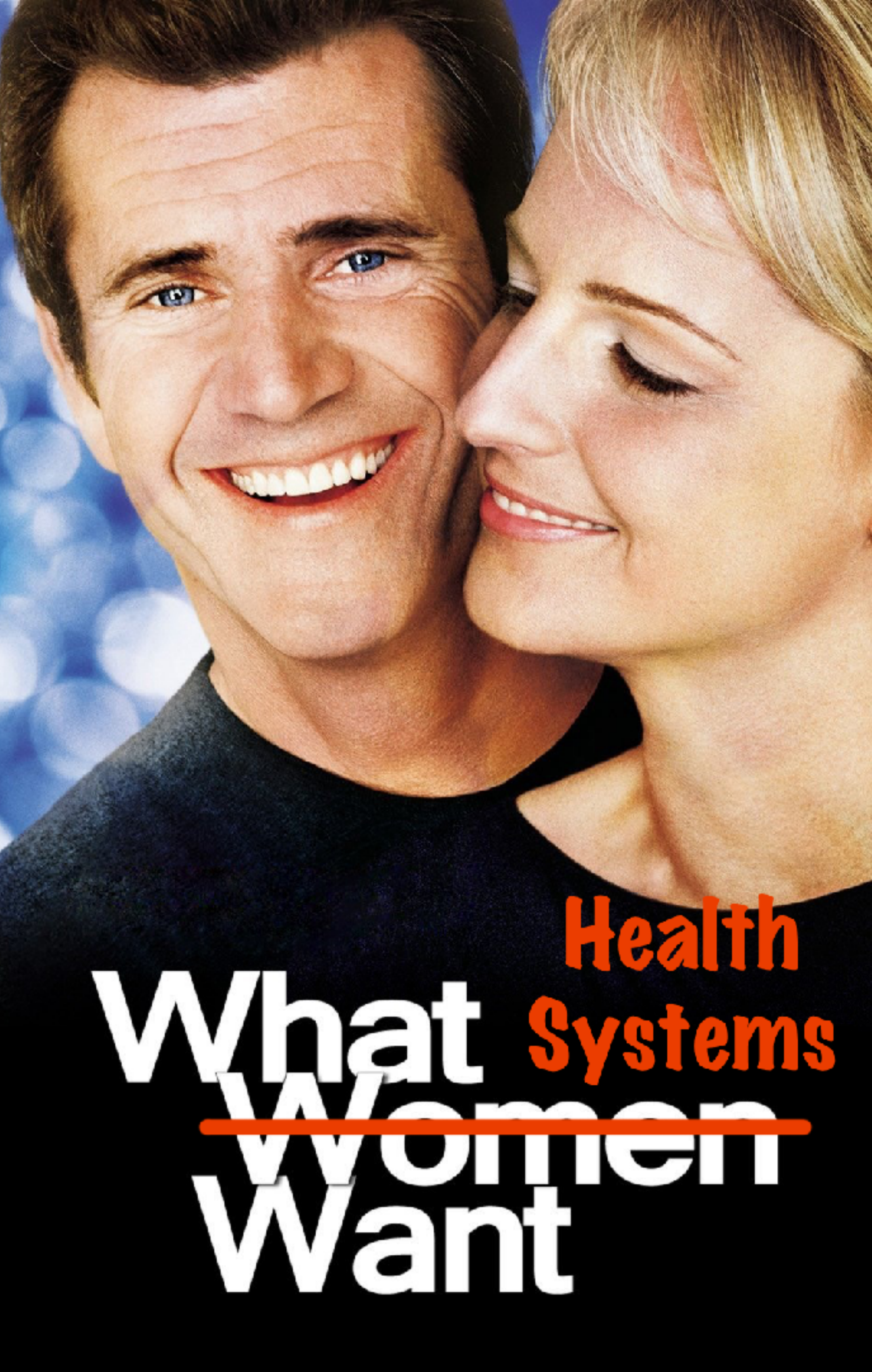
Healthcare Today



THE REALITY

		Score
L	Length of Stay (days)	1
		2
		3
		4-6
		7-13
		≥ 14
A	Acuity of Admission	Elective
		Urgent/Emergent
C	Charlson Comorbidity Score	0
		1
		2
		3
		≥ 4
E	Emergency Department Visits in the Preceding Six Months	0
		1
		2
		3
		≥ 4

Add component scores from boxes L, A, C, E to calculate total LACE score.



What Health Systems
~~Women~~ Want

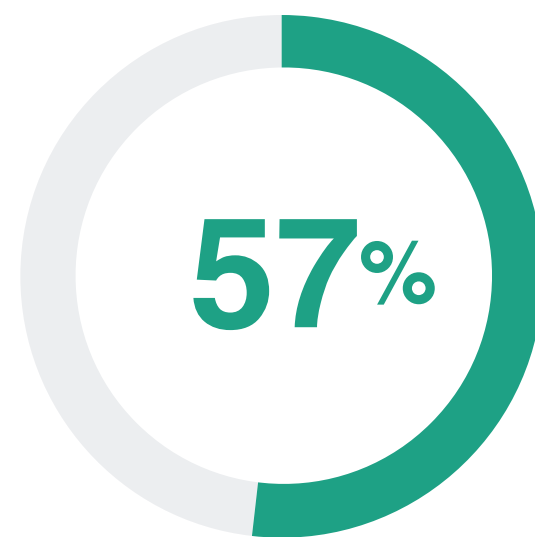


Where ML is Needed



It's all about outcomes

What are the most valuable outcomes to forecast with predictive analytics? Providers weigh in.



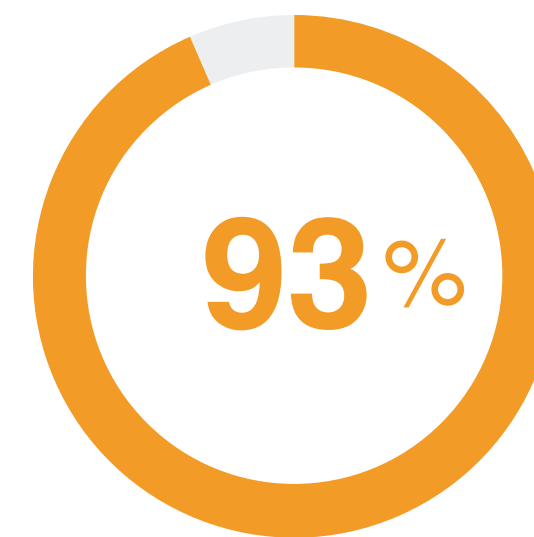
Savings

57% of executives forecast predictive analytics will save their organization 15% or more over the next 5 years



Currently Use

89% of providers currently use predictive analytics or plan to begin in the next 5 years



Importance

93% say predictive analytics is important to the future of their business

Most Valuable Outcomes to Predict Using Predictive Analytics

53% Patient Satisfaction

38% Patient Demand

48% Hospital Readmissions

37% Profitability

46% Staffing and Workforce Needs

36% Cost

41% Clinical Outcomes

34% Diagnosis

Biggest Challenges to Implementation Within Organizations

16% Lack of Budget

10% Lack of Sufficient Tech

13% Regulatory Issues

9% Too Much Data

12% Incomplete Data

8% Patient Matching

11% Lack of Skilled Employees

7% Lack of Confidence



What is healthcare.ai?



And how can it help?

healthcare.ai



Community

Growing website and online community focused on healthcare machine learning.



Education

Weekly blogs and YouTube live event broadcasts where you can engage the team with questions and get answers.



Open Source

Open source R and Python packages are designed to streamline healthcare machine learning

Who is it for?



Software Developers



**BI Developers, Analysts, &
Data Architects**



Data Scientists

Machine Learning Packages

Healthcare.ai is available in packages for both R and Python, two of the most common languages used by data scientists. If you don't have previous experience with either language, we recommend the R package as it currently has more features and is more user-friendly.



R Package

<https://github.com/HealthCatalyst/healthcareai-r>



Python Package

<https://github.com/HealthCatalyst/healthcareai-py>

Features

What does it do?



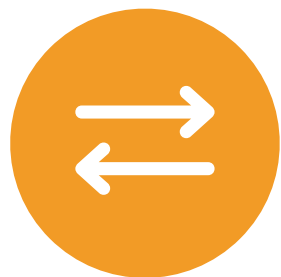
Analyze data using statistical methods that are relevant to healthcare

Provides statistical tools that are useful for analyzing healthcare data, pays attention to longitudinal questions, and offers an easy way to do risk-adjusted comparisons



Create, save, and deploy machine learning models

Makes it easy to load data and create models from various data sources and then save and deploy the models so they can be used to make predictions



Easily compare and evaluate machine learning models

Provides appropriate metrics to assess which algorithm generates the best model and tells you which features are most important to your model

Healthcare Statistics

Both packages differ from other machine learning packages in that they focus on data issues specific to healthcare. This means that it pays attention to longitudinal questions, offers an easy way to do risk-adjusted comparisons, and provide easy connections and deployment to databases.



Find high variation

Search across subgroups and measures (LOS, readmissions) to find groups that need improvement



Calculate All Correlations

View the correlations between all variables in your dataset



Find Trends

See if columns have been trending up or down over time using automatic subgroups like Gender



Risk-adjusted comparisons

Compare healthcare groups on a particular measure (like mortality rate), adjusting for the health of the patients



Calculate Target Correlations

View the correlations between several variables and your variable of interest

Find Variance Example

```

1 library(healthcareai)
2 df <- data.frame(Dept = c('A','A','A','B','B','B','B','B'),
3   Gender = c('F','M','M','M','M','F','F','F'),
4   LOS = c(3.2,NA,5,1.3,2.4,4,9,10))
5
6 categoricalCols <- c("Dept","Gender")
7
8 dfRes <- findVariation(df = df,
9   categoricalCols = categoricalCols,
10  measureColumn = "LOS")
11
12 dfRes

```

DimensionalAttributes	CategoriesGrouped	MeasureCOV	MeasureVolumeRaw	MeasureVolumePercent	MeasureImpact	AboveMeanCOVFLG	AboveMeanVolumeFLG
Dept	A	LOS 1.73	LOS 3	LOS 0.38	LOS 5.19	LOS Y	LOS N
Dept	B	LOS 0.91	LOS 5	LOS 0.62	LOS 4.55	LOS N	LOS Y
Gender	F	LOS 1.15	LOS 4	LOS 0.5	LOS 4.6	LOS N	LOS N
Gender	M	LOS 1.15	LOS 4	LOS 0.5	LOS 4.6	LOS N	LOS N
Dept Gender	A M	LOS 1.41	LOS 2	LOS 0.29	LOS 2.82	LOS Y	LOS N
Dept Gender	B M	LOS 1.41	LOS 2	LOS 0.29	LOS 2.82	LOS Y	LOS N
Dept Gender	B F	LOS 0.87	LOS 3	LOS 0.43	LOS 2.61	LOS N	LOS Y

Streamlined Workflow

Both packages offer a streamlined workflow that make it easy to create, train, evaluate, save, and deploy machine learning models. The algorithms and techniques used within both packages are suitable for healthcare data.



Supervised Model Trainer

Streamlined trainer with pre-selecte defaults. Suitable entry point for beginners because it abstracts away challenging parts of training models.



Advanced Model Trainer

Advanced trainer for users that may want to use different data preparation pipelines or tune hyperparameters.



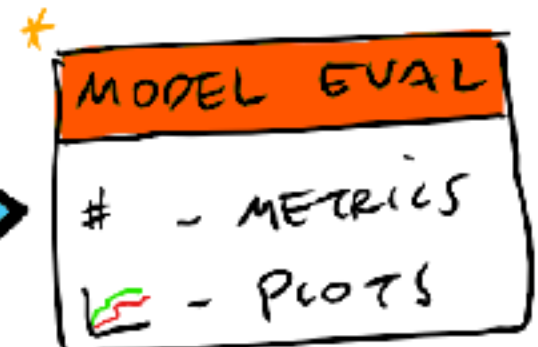
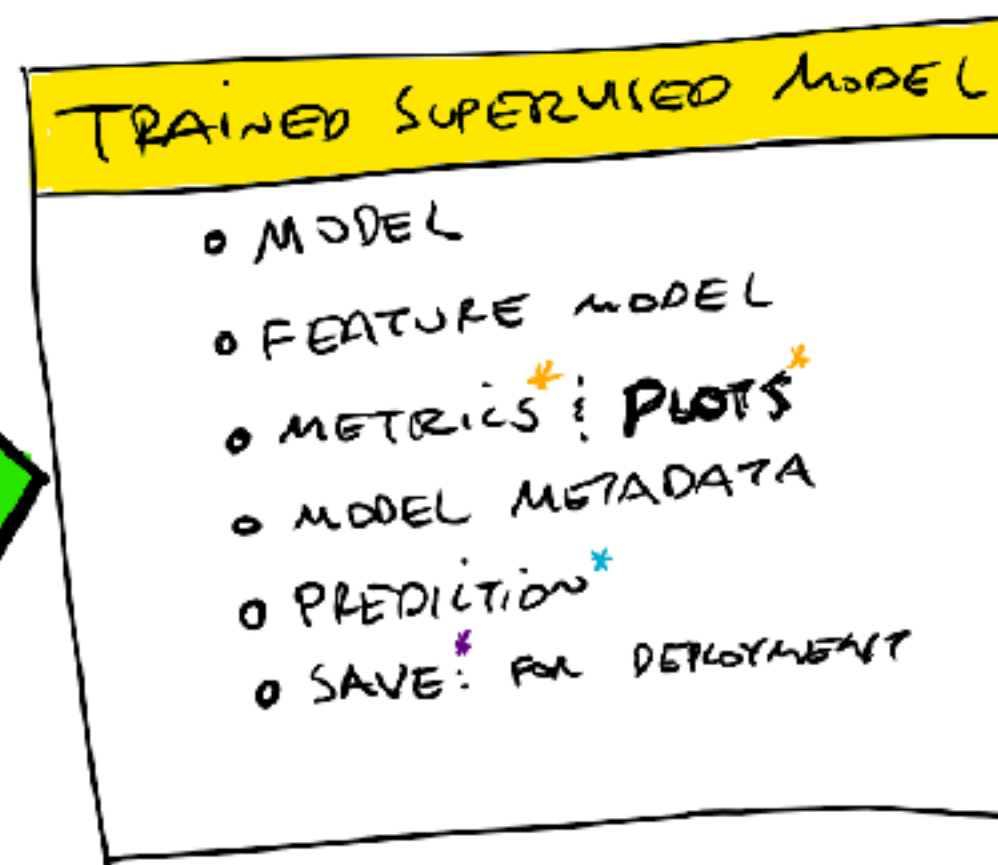
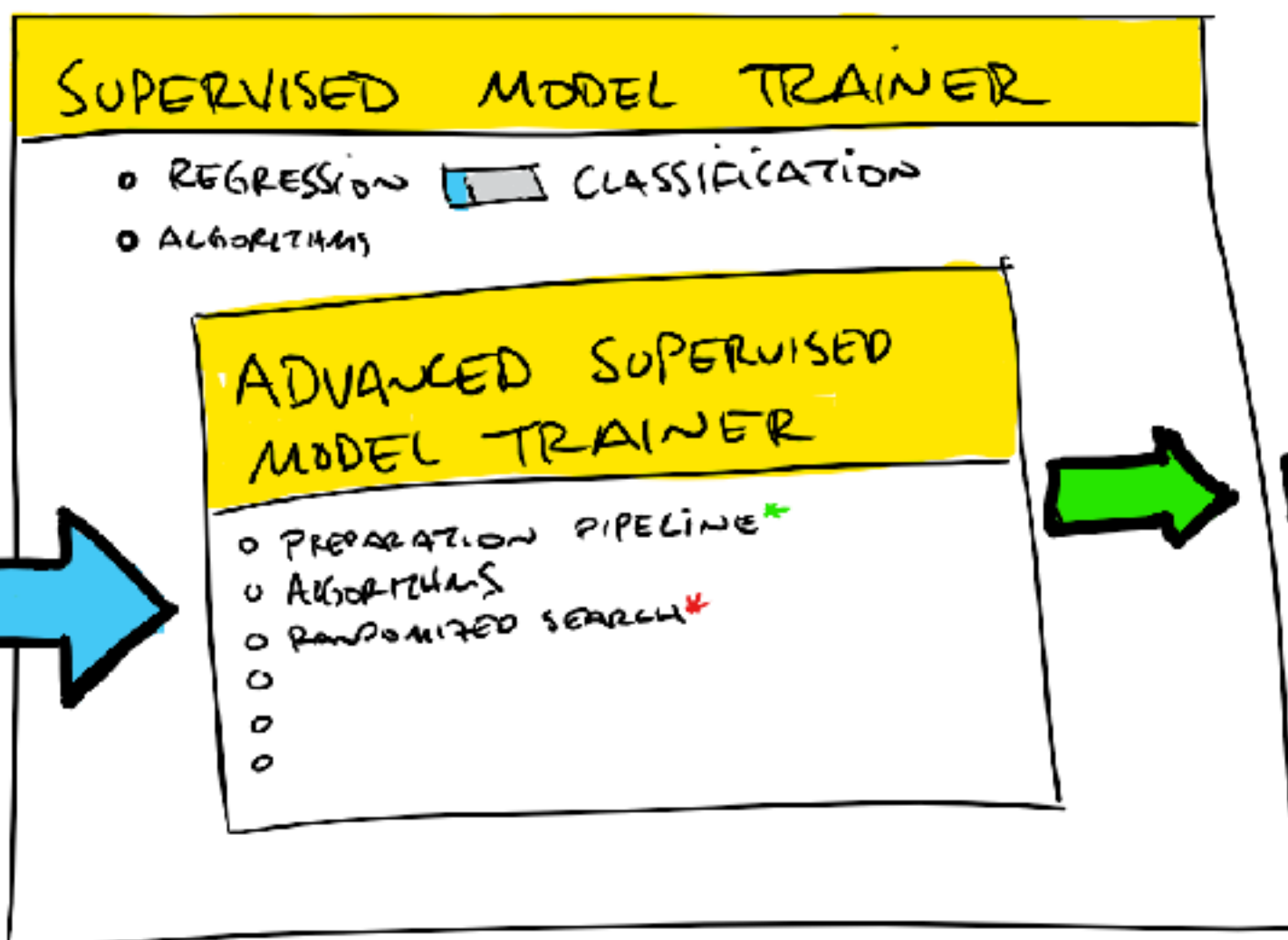
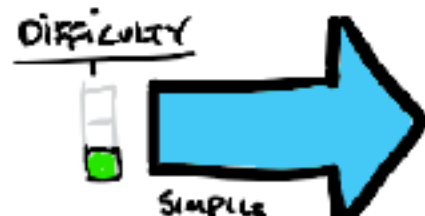
Trained Supervised Model

Returned by trainer classes and contains metadata, metrics, plots, ans save methods.



HEALTHCARE-AI

DEVELOPER ARCHITECTURE OVERVIEW



- * FILTERS
- * TRANSFORMERS
- * RANDOMIZED SEARCH
- TOP FACTORS

- * WRITE TO DATABASES
- * FILE IO UTILITIES

CLASSES

MODULES

Create, Train, Save Example

```
1 import healthcareai
2
3 # Load training and prediction data sets
4 training_df = healthcareai.load_diabetes()
5 prediction_df = healthcareai.load_diabetes()
6
7 # Create a model trainer and train it
8 trainer = healthcareai.SupervisedModelTrainer(
9     dataframe=training_df,
10    model_type='classification',
11    predicted_column='ThirtyDayReadmitFLG',
12    grain_column='PatientEncounterID')
13
14 # Train a Logistic Regression model
15 trained_model = trainer.logistic_regression()
16 trained_model.save(filename='healthcare_ai_model.pkl')
17
18 # Make predictions and save them
19 predictions = trained_model.make_predictions(prediction_df)
20 predictions.to_csv('healthcare_ai_predictions.csv')
21
```


Regression

Classification

Listed by increasing sophistication

Linear Model

Lasso \mathbb{R} \oplus

Ridge \mathbb{R}^* \oplus

Random Forest \mathbb{R} \oplus

k-Nearest Neighbors \oplus

Linear Mixed Model \mathbb{R}

XGBoost \mathbb{R}^* \oplus^*

Neural Network \oplus

Linear Model

Lasso \mathbb{R} \oplus

Ridge \mathbb{R}^* \oplus

Random Forest \mathbb{R} \oplus

k-Nearest Neighbors \oplus

Support Vector Machine \oplus

Linear Mixed Model \mathbb{R}

XGBoost \mathbb{R} \oplus^*

Neural Network \oplus

Compare within categories

Find Variation \mathbb{R}

Visualize difference between groups

Variation Across Groups \mathbb{R}

Clustering

K-means \mathbb{R}

Need to reduce dimensions?



Dimension Reduction

Principle Component Analysis \mathbb{R}



DEMO



Readmissions Example



Real World Use Cases



How is it being used in the wild?

Metrics

- Saving > \$1M annually predicting financial codes at a major health system
- Reducing central-line infection (CLABSI) rates by >50% at a large academic medical center
- Lowering patient financial burden across 150k patients per month at large midwest health system
- Helping clinicians and operations to prioritize resources with predictions from models with literature-beating accuracy across readmissions, infection, and finance.



Questions?