



Quantifying Dynamic Risk with Predictive Analytics (AR10)

PhUSE Connect 2019

Susan Olson, Ph.D. and Connor English, MPH

EdjAnalytics

Today's Presentation

- Data Science Overview
- Predictive Modeling Workflow
- Dynamic Risk Application 1 – Hospital Readmission
- Dynamic Risk Application 2 – Elective Procedures
- Closing Remarks

Data Science Overview

Data Science Overview

- **Data science helps organizations drive decision making with insights from data.**
- Using research, data engineering, machine learning, and other advanced technologies, data science teams engineer actionable, accurate workflow solutions from large sets of data.
- From life sciences to financial services to sports to education, the same approach and technological methods serve a spectrum of industries and organizations.



Analytics Approach

- Data science can use a formulaic approach, stepping up from assessing data to predicting and prescribing solutions in any application.

1. ASSESSMENT

What information do we have?

2. DESCRIPTIVE

What already happened?

3. PREDICTIVE

What's likely to happen?

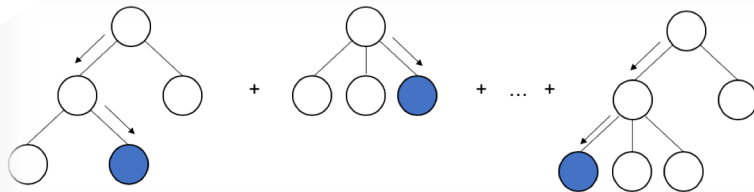
4. PRESCRIPTIVE

What should we do next?

- The available data and its quality determine what next steps are possible.

Sample Analytics Methods

- Deep learning, convolutional neural networks, temporal convolutional networks, gradient boosting machine, support vector machine, random forest, multiple logistic regression, decision tree classification and others
- Bayesian networks, probability modeling and naïve Bayes classifiers
- K-means clustering, K-nearest-neighbors clustering, and principal component analysis
- Monte Carlo simulations
- Standard statistical testing: t-tests, chi-squared tests, ANOVAs, and others
- Superiority testing
- Diagnostic testing & bootstrapping (sampling with replacement) for confidence interval generation
- Survival analysis, Cox models, and non-parametric testing
- Survey design and analysis



Sample Solutions

Segmenting Populations



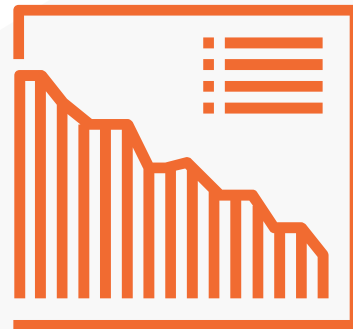
What clusters are observed in patient populations?

Dynamic Risk/Opportunity Assessment



As a function of time, what patients are at future risk for adverse events? Are there opportunities for improved outcomes?

Ranking & Queuing



What risk indicators should be prioritized? What risk factors are most effectively mitigated?

Predictive Modeling Workflow

Locating and Selecting Data

- Identify historical data and outcomes to use as the basis of the predictive model.
- Locate the data sources and the stewards/curators of those data sources.
- Consult organizational practices, experts and rules regarding privacy and security.
- Identify the database, data flow, computing environment and technology stack to use for analytics.

Consider a variety of data sources for predictive modeling. Predictive value may be discovered in non-intuitive features.



Data Assessment

Assess Readiness for Predictive Modeling

Data Inventory

- ✓ Variable identification (e.g. name, type)
- ✓ Structural nature of data (e.g. may be used directly vs. needs transformation)
- ✓ Data missingness within variable set

Data Identification

- ✓ Outcome interpretation
- ✓ Number of examples within data set per outcome
- ✓ Useable set of variables for analysis

Data Interpretation

- ✓ Meaning of independent variables
- ✓ Preparation for feature engineering
- ✓ Analytic technique assessment

Data Cleaning & Feature Engineering

Cleaning example

Patient ID	Weight (lb)
607584	153
607593	203
608365	795
608218	193
...	...

Feature engineering example

Patient ID	Smoking status	Smoking status (Engr)
607584	Non	0
607593	Occasional	1
608365	Frequent	1
608218	Former	0
...	...	

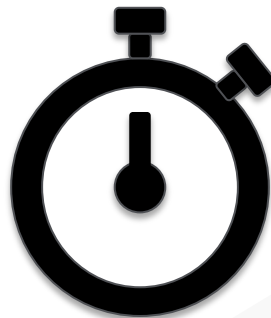
- Data aggregation and imputation processes follow.

Training, Validation & Testing Data Sets



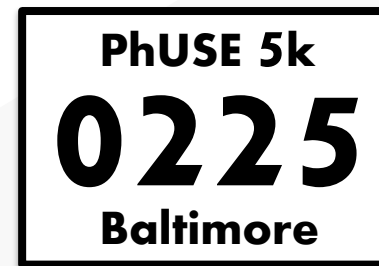
Training set

70 – 80% of full data



Validation set

10 – 15% of full data

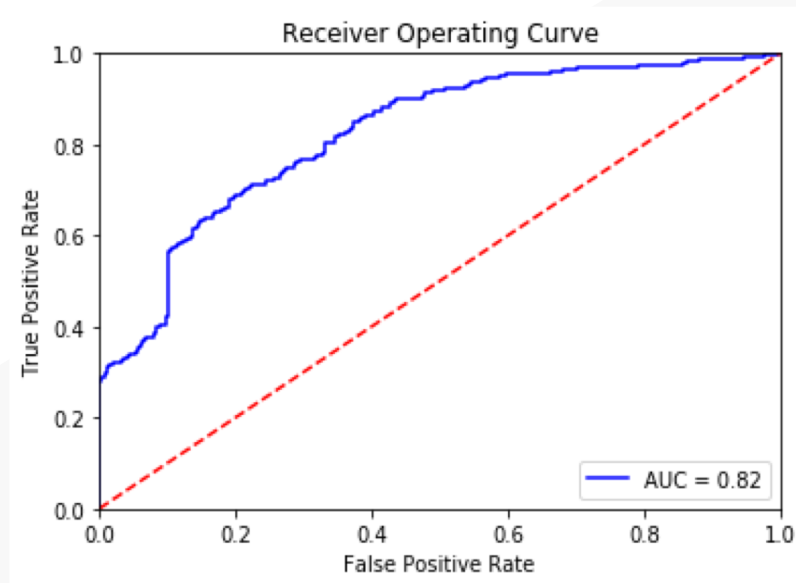


Testing set

10 – 15% of full data

Predictive Model Execution & Performance Testing

- Model execution is iterative; feature engineering may be revisited multiple times.
- Multiple model types may be run to discover best fit for performance.
- Hyperparameters (static properties of the model set before training such as how fast it should learn or how many rounds of training should occur) are tuned using the validation set predictions.
- Once the model is performing satisfactorily with the validation set, performance testing occurs.



**The closer to 1 the AUC, the more accurate the model's predictions are.*

Dynamic Risk Application 1 - Hospital Readmission

Goal, Data & Approach

- **Goal**

- Decrease 30-day readmission rates of CHF patients covered by Medicare for tri-state hospital system with 900 beds, 42 locations and 9,500 employees

- **Data**

- Three years of medical history (including EMR data)
- Observations of unplanned readmissions
- Demographics for CHF patients covered by Medicare

- **Approach**

- Predictive model dynamically quantifying **each patient's risk** of readmission

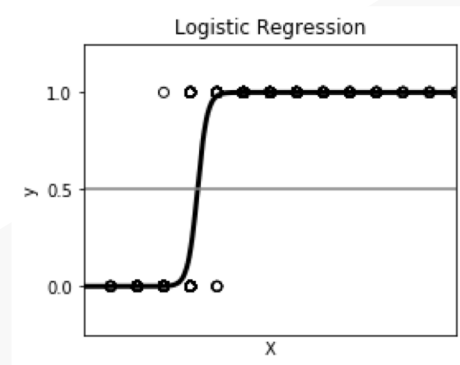
Model Approach & Results

- **Model approach**

- Logistic regression
- Engineered feature examples
 - Summary vital measures
 - Defined comorbidities
 - Previous visit indicators
 - Discharge disposition

- **Results**

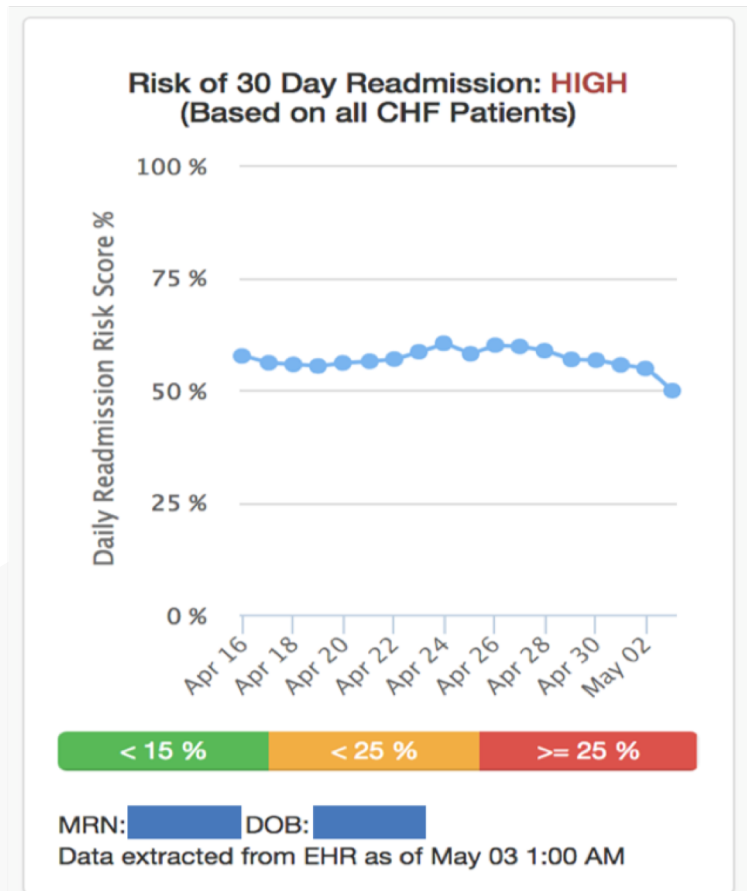
- Variables of importance
 - Type of facility where a patient is discharged
 - Number of previous encounters where CHF was listed as a co-morbidity
 - Number of times the patient visited the hospital in the previous 90 days
 - Potassium, Hemoglobin and Creatinine levels
- AUC = 0.7



Dynamic Risk Assessment

- **Monitoring**

- Daily Risk Score %
 - Real-time analysis
 - In the example at right, there is a 50% risk of readmission based on the latest information. Conditions improved from the previous marker.
 - Risk can be categorized into H/M/L to facilitate data visualization.



Dynamic Risk Assessment

- **Discharge Planning**
 - Discharge Disposition
 - User selects planned discharge disposition
 - Relative risk of readmission shown for selection
 - Risk may be color coded for data visualization

Relative Risk of 30 Day Readmission: **HIGH** (Based on Planned Discharge Disposition)

If you have a planned discharge disposition for this patient, select the discharge disposition below to see how the risk for 30-day readmission for this patient compares with other patients in the selected discharge cohort.

Home with Home Care Services



This patient is in the **99th Percentile** of other patients who had this discharge disposition.

Dynamic Risk Application 2 – Elective Procedures

Background, Data & Approach

- **Goal**

- Identify which patients are most likely to undergo elective surgeries
 - Hip Replacement
 - Knee Replacement
 - Bariatric Surgery
- Improve outreach efforts to candidates about health benefits
- Inform clinicians of at-risk patients

- **Data**

- 7.7 million total procedures (EMR data), five years of history

- **Approach**

- Extreme Gradient Boosting model (XGBoost)



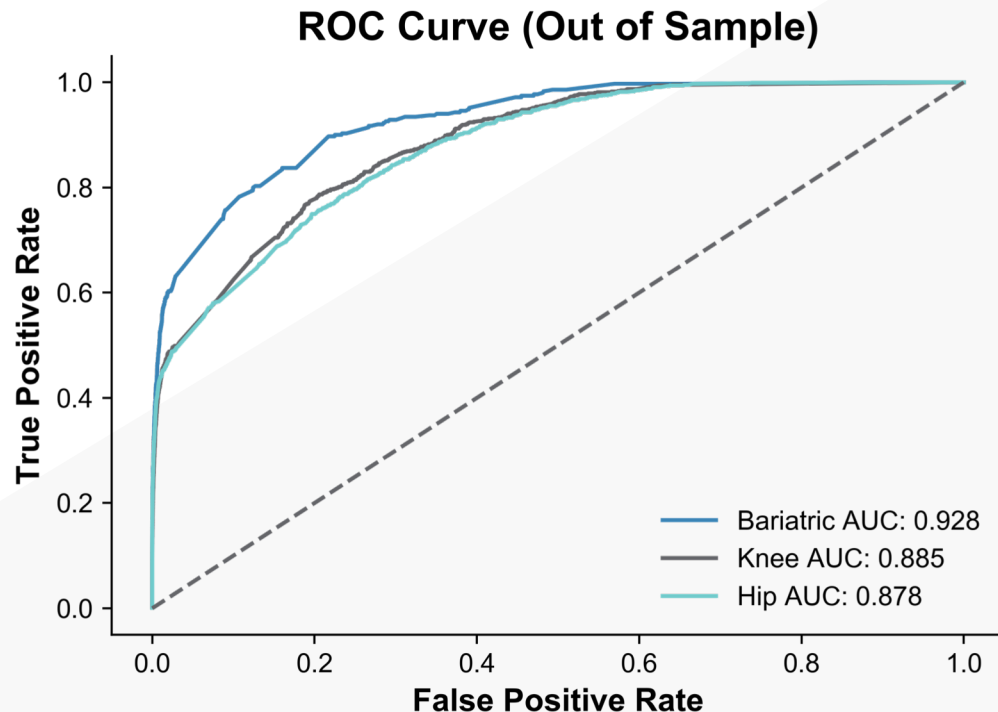
Model Approach & Results

- **Data Preparation**

- Data aggregation was key to accurately predicting dynamic risk to assure the model did not include knowledge of the future.
- About 80 features included

- **Results**

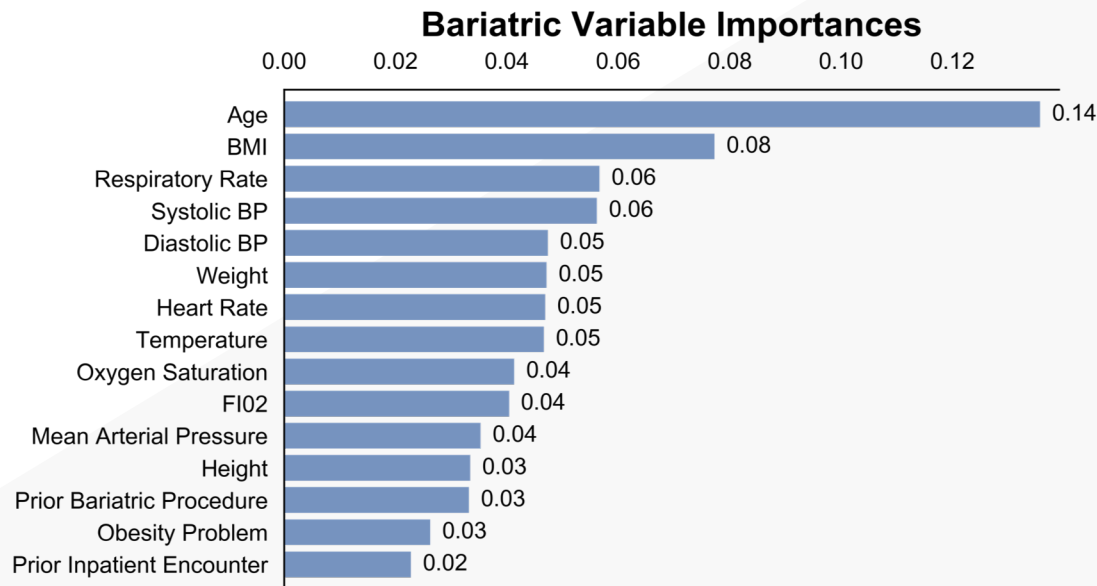
- Delivered 3 models with strong AUCs
- Risk scores translated into High, Medium, and Low levels



Model Results: Variables of Importance

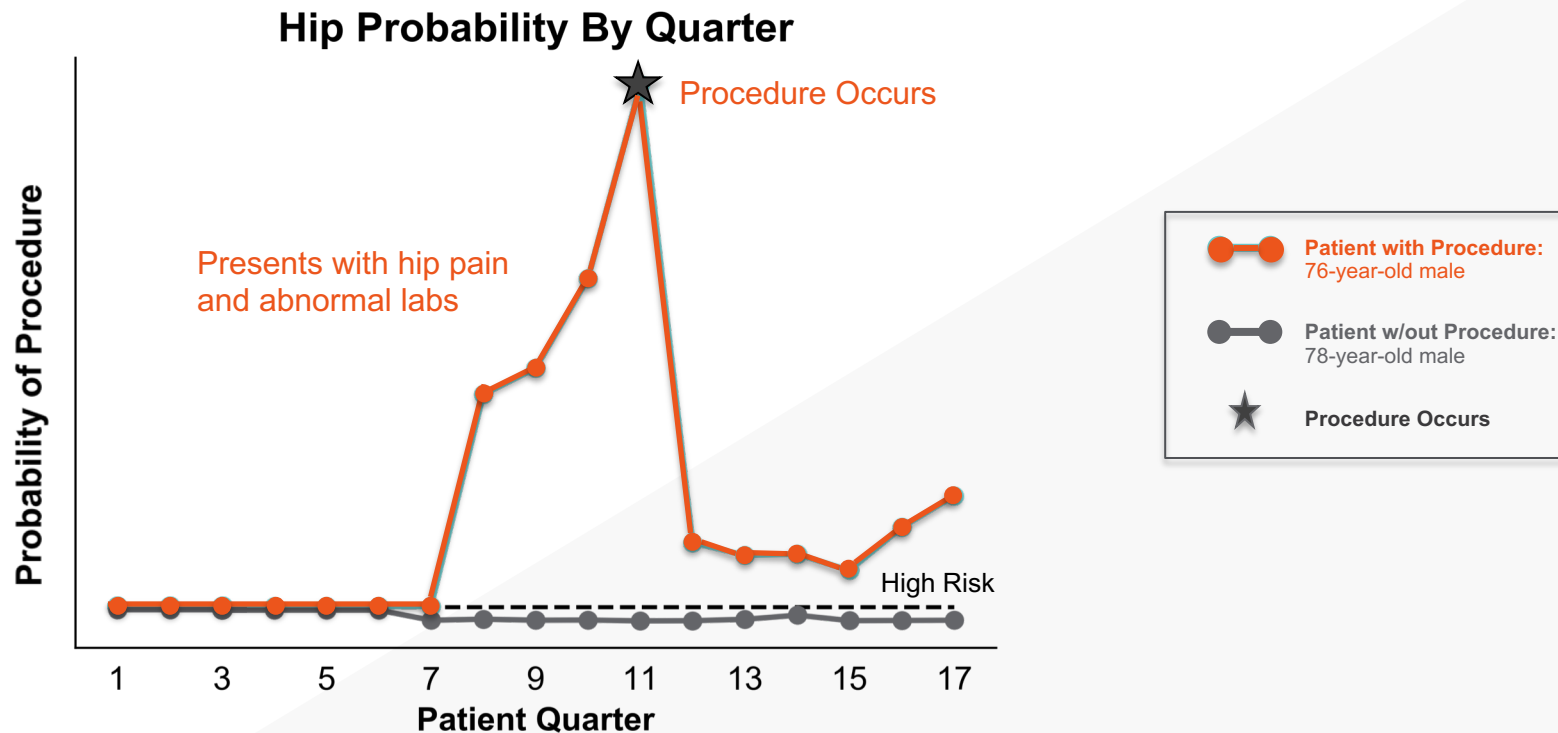
- Findings**

- Age and BMI rank high in the variables of importance.
- Substantially more accurate results come from considering many other factors, including vitals and prior encounters.



Features ranked by relative contribution to model accuracy

Example Patient-Level Dynamic Risk View



Example Aggregations by Risk Level and Quarter

- **Sample Output**

Knee Patients - Risk Classification			
Quarter	High Risk	Medium Risk	Low Risk
Q1	11,740	54,034	78,344
Q2	12,092	55,414	75,802
Q3	11,549	56,382	79,943
Q4	11,842	58,681	83,429
Unique Patients across Quarters	32,867	91,592	153,842

Closing Remarks

Closing Remarks

- The analytics methods presented today serve a spectrum of industries and are foundational to many applications.
- Other predictive analytics concepts include:
 - Adverse Events Detection, Drug Pipeline Efficiencies, Public Health Impact/Audit Risk, Drug Shortages, Social Determinants & Drug Needs, Value-based Contracting and Evidence-based Outcomes



Contact Us

Susan Olson

Chief Operating Officer

solson@edjanalytics.com

(574) 210-6592

Connor English

Data Scientist

cenglish@edjanalytics.com

(502) 208-4088