UNIVERSITY of MARYLAND
MEDICAL SYSTEM

Machine Learning in the Age of Healthcare Analytics

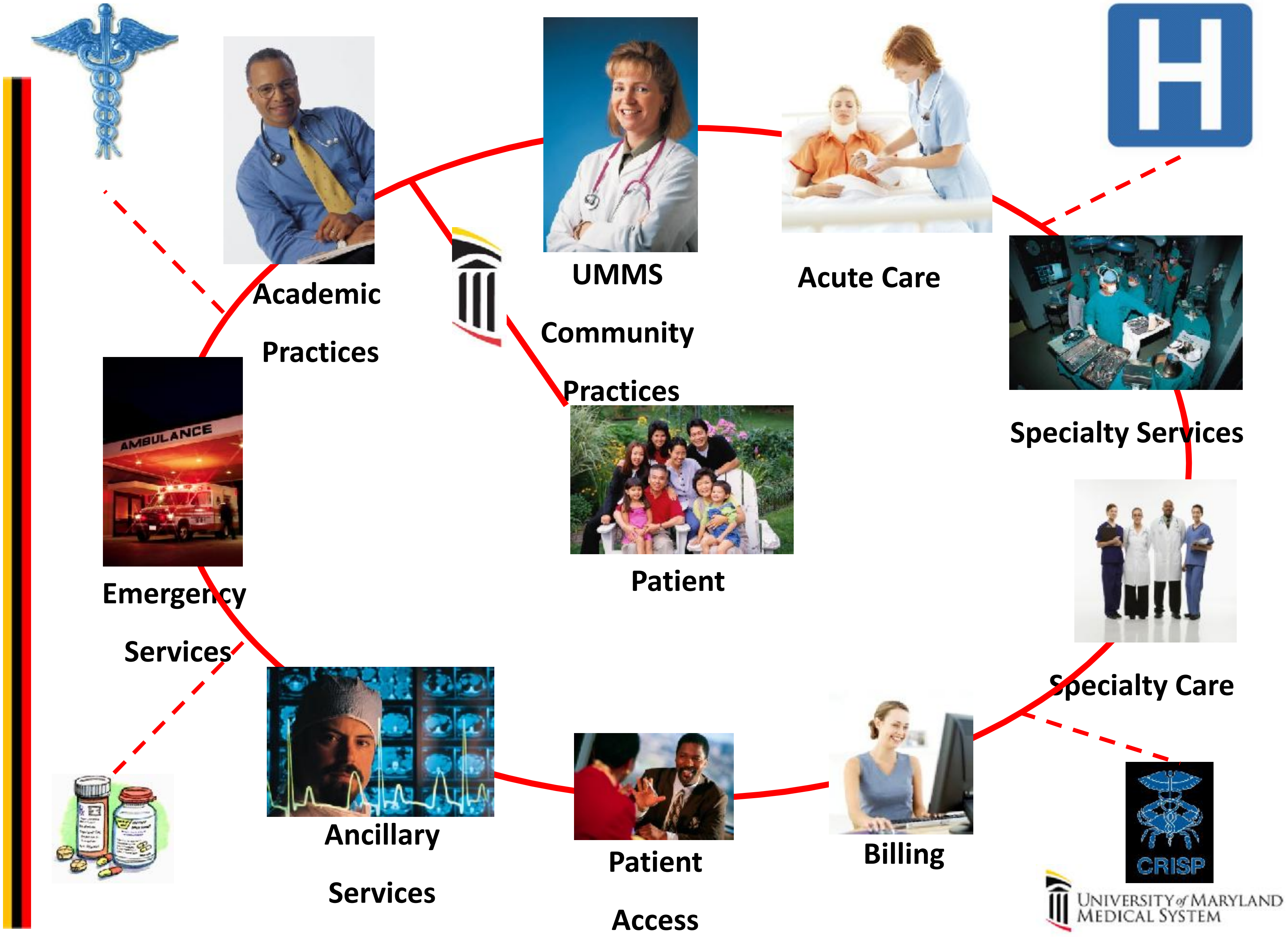
Warren D'Souza, PhD, MBA, FAAPM

Vice President, Enterprise Data & Analytics

University of Maryland Medical System

Healthcare Ecosystem

One Patient – One Record – One Portfolio



Academic Practices

UMMS Community Practices

Acute Care

Specialty Services

Specialty Care

Billing

Patient Access

Ancillary Services

Emergency Services

Patient

UNIVERSITY of MARYLAND MEDICAL SYSTEM

Value-based Healthcare

- Transition from *fee-for-service* to *value-based care*
  - Rising health care costs
  - Clinical inefficiency
  - Duplication of services

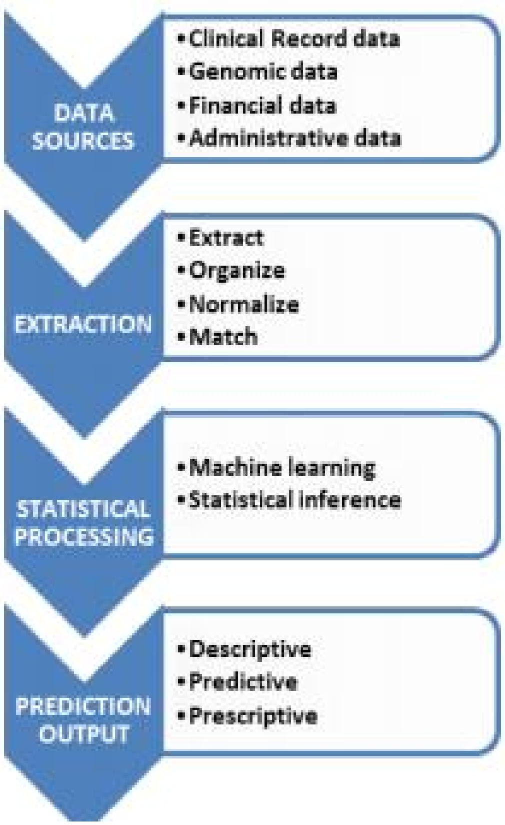
$$\text{Value} = \frac{\text{Health outcomes}}{\text{Costs of delivering the outcomes}}$$

M. Porter - HBS

- Value-based payment models
  - Accountable Care Organization
  - Patient-Centered Medical Homes
  - Pay-for-Performance
  - Bundled Payments

Healthcare Analytics

- ## Healthcare “Big Data”
  - Volume – ever increasing amounts
  - Velocity – quickly generated
  - Variety – many different types
  - Veracity – from trustable sources



DATA SOURCES
- Clinical Record data
- Genomic data
- Financial data
- Administrative data

EXTRACTION
- Extract
- Organize
- Normalize
- Match

STATISTICAL PROCESSING
- Machine learning
- Statistical inference

PREDICTION OUTPUT
- Descriptive
- Predictive
- Prescriptive

The Analytics Pipeline

Source: Kumar et al, Comm. of the ACM, 2013

UNIVERSITY of MARYLAND MEDICAL SYSTEM

systems

technology

engineering

historical

data

patterns

action

models

process

outcomes

Clinical

Probabilistic

people

optimize

Analytics

quantitative

reports

decisions

Predictive

qualitative

statistics

quality

Operations

research

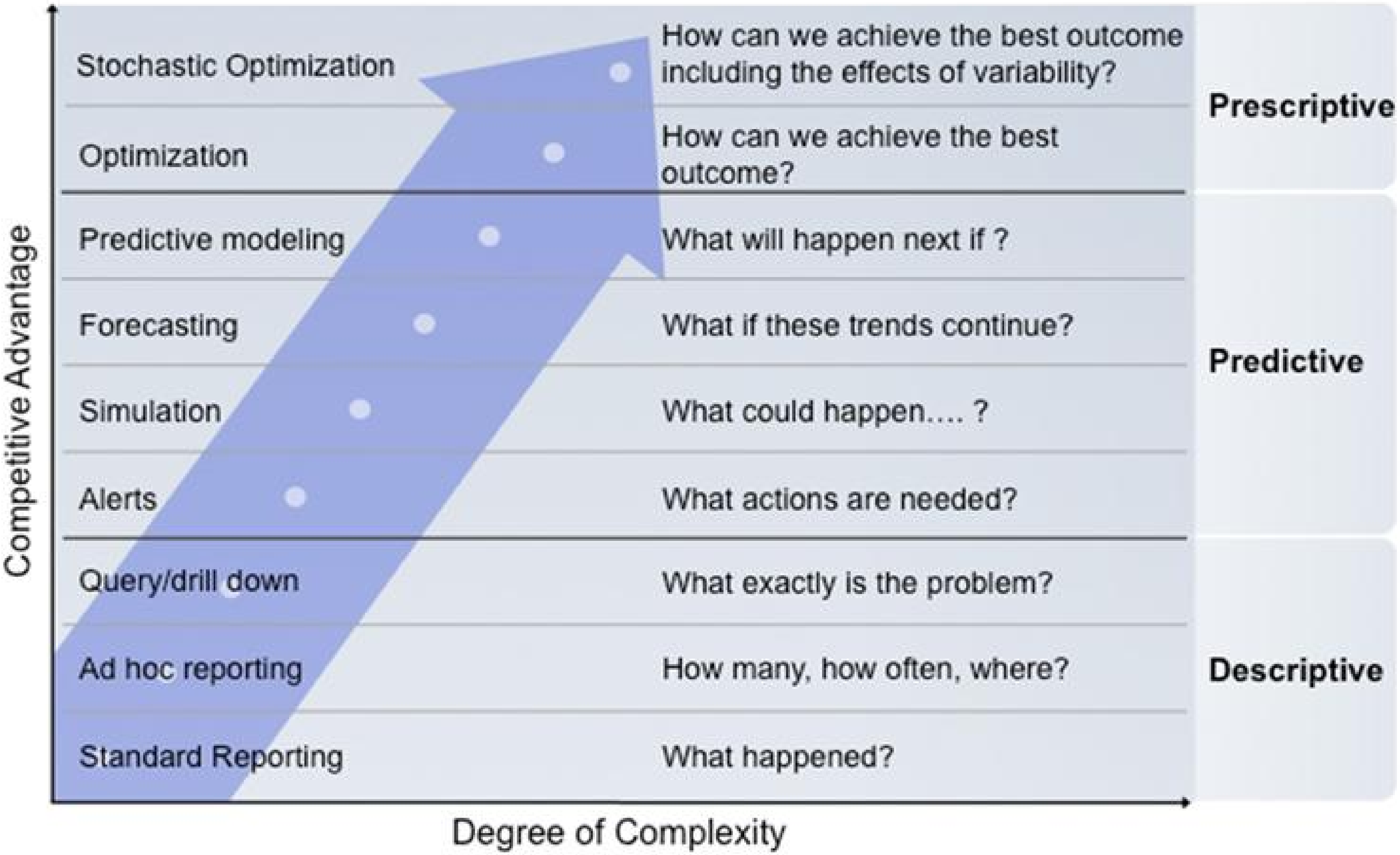
Finance

performance

analysis

standards

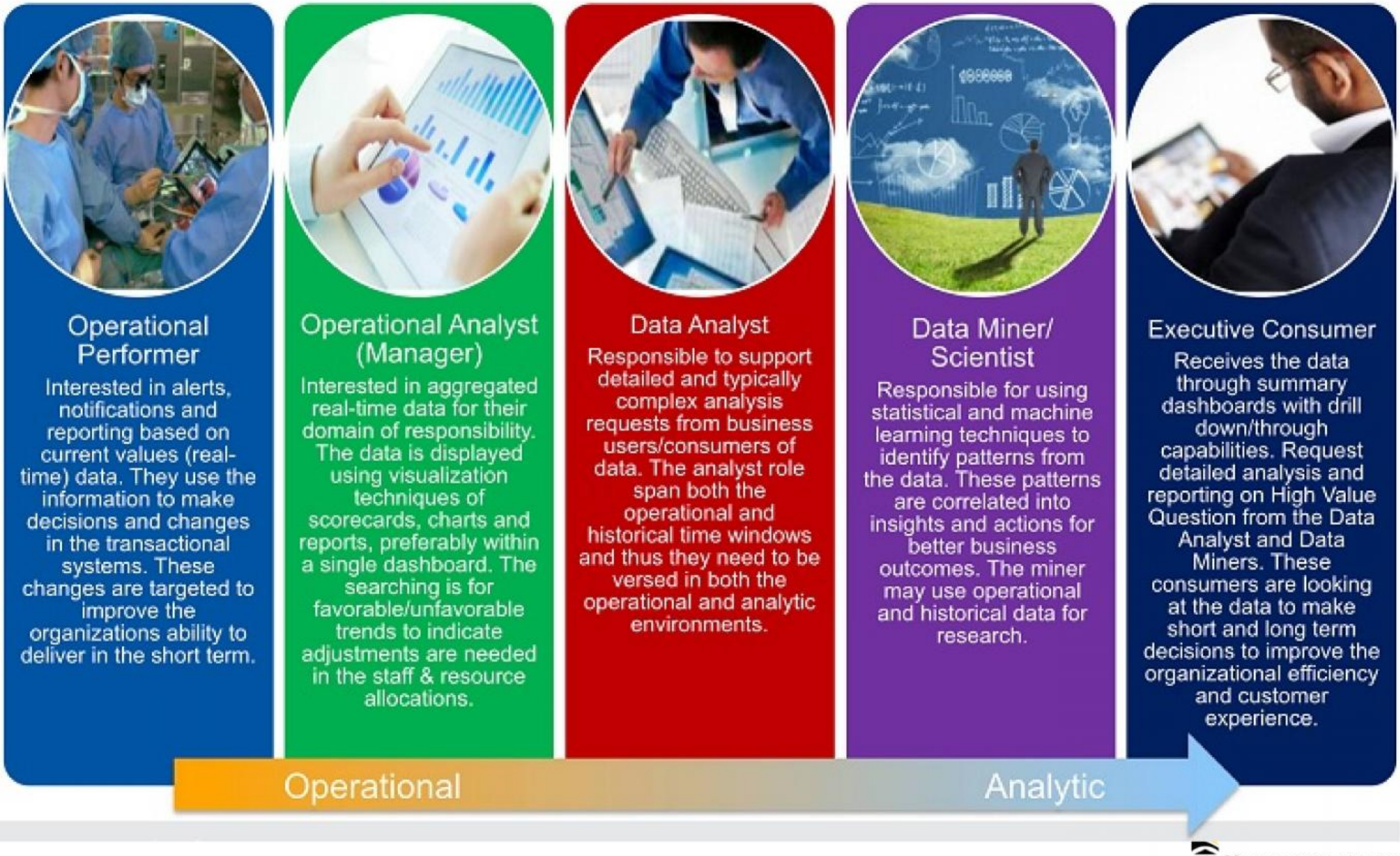
What is Analytics?



Stochastic Optimization	How can we achieve the best outcome including the effects of variability?	Prescriptive
Optimization	How can we achieve the best outcome?	
Predictive modeling	What will happen next if ?	Predictive
Forecasting	What if these trends continue?	
Simulation	What could happen.... ?	
Alerts	What actions are needed?	
Query/drill down	What exactly is the problem?	Descriptive
Ad hoc reporting	How many, how often, where?	
Standard Reporting	What happened?	

Based on: Competing on Analytics, Davenport and Harris, 2007

Data Personas (The Requirements)



Operational Performer

Interested in alerts, notifications and reporting based on current values (real-time) data. They use the information to make decisions and changes in the transactional systems. These changes are targeted to improve the organizations ability to deliver in the short term.

Operational Analyst (Manager)

Interested in aggregated real-time data for their domain of responsibility. The data is displayed using visualization techniques of scorecards, charts and reports, preferably within a single dashboard. The searching is for favorable/unfavorable trends to indicate adjustments are needed in the staff & resource allocations.

Data Analyst

Responsible to support detailed and typically complex analysis requests from business users/consumers of data. The analyst role span both the operational and historical time windows and thus they need to be versed in both the operational and analytic environments.

Data Miner/ Scientist

Responsible for using statistical and machine learning techniques to identify patterns from the data. These patterns are correlated into insights and actions for better business outcomes. The miner may use operational and historical data for research.

Executive Consumer

Receives the data through summary dashboards with drill down/through capabilities. Request detailed analysis and reporting on High Value Question from the Data Analyst and Data Miners. These consumers are looking at the data to make short and long term decisions to improve the organizational efficiency and customer experience.

Transforming Medicine with Algorithms

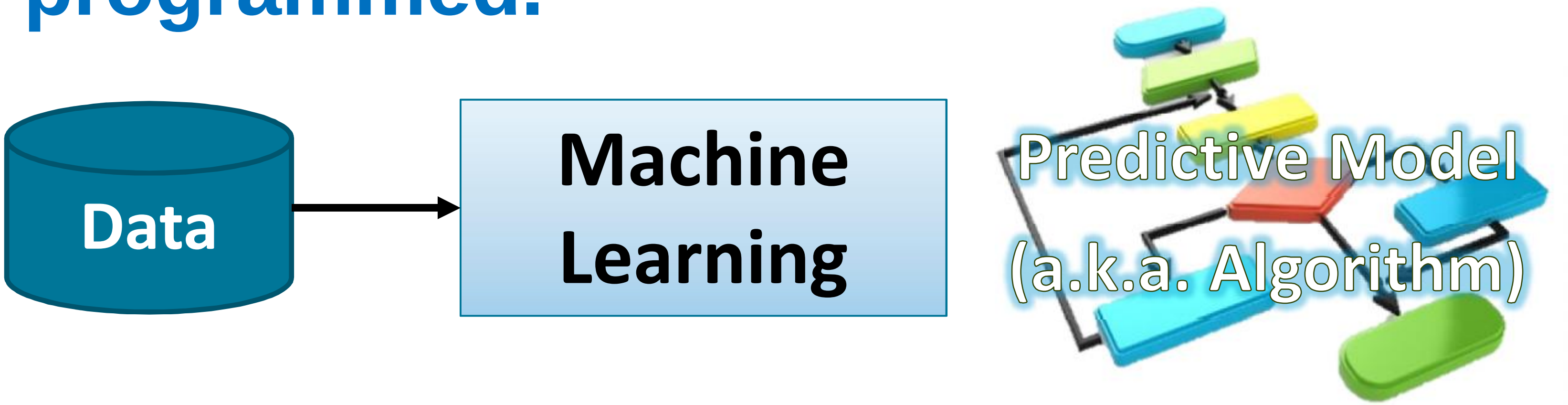
- **Big Data will transform medicine**
  - **EHRs**
  - **Claims**
  - **Socio-economic**
  - **Social**

- **Algorithms acting on data will prove transformative**
  - **Shift from traditional statistical approaches to newer computational approaches**
  - **Ability to handle large numbers of observations and (outcome) predictor variables**

What is Machine Learning?

Machine Learning is ……..

a field that provides computers with the ability to learn (from observations and experience) without being explicitly programmed.



Data

Machine Learning

Predictive Model

(a.k.a. Algorithm)

Origins of Machine Learning

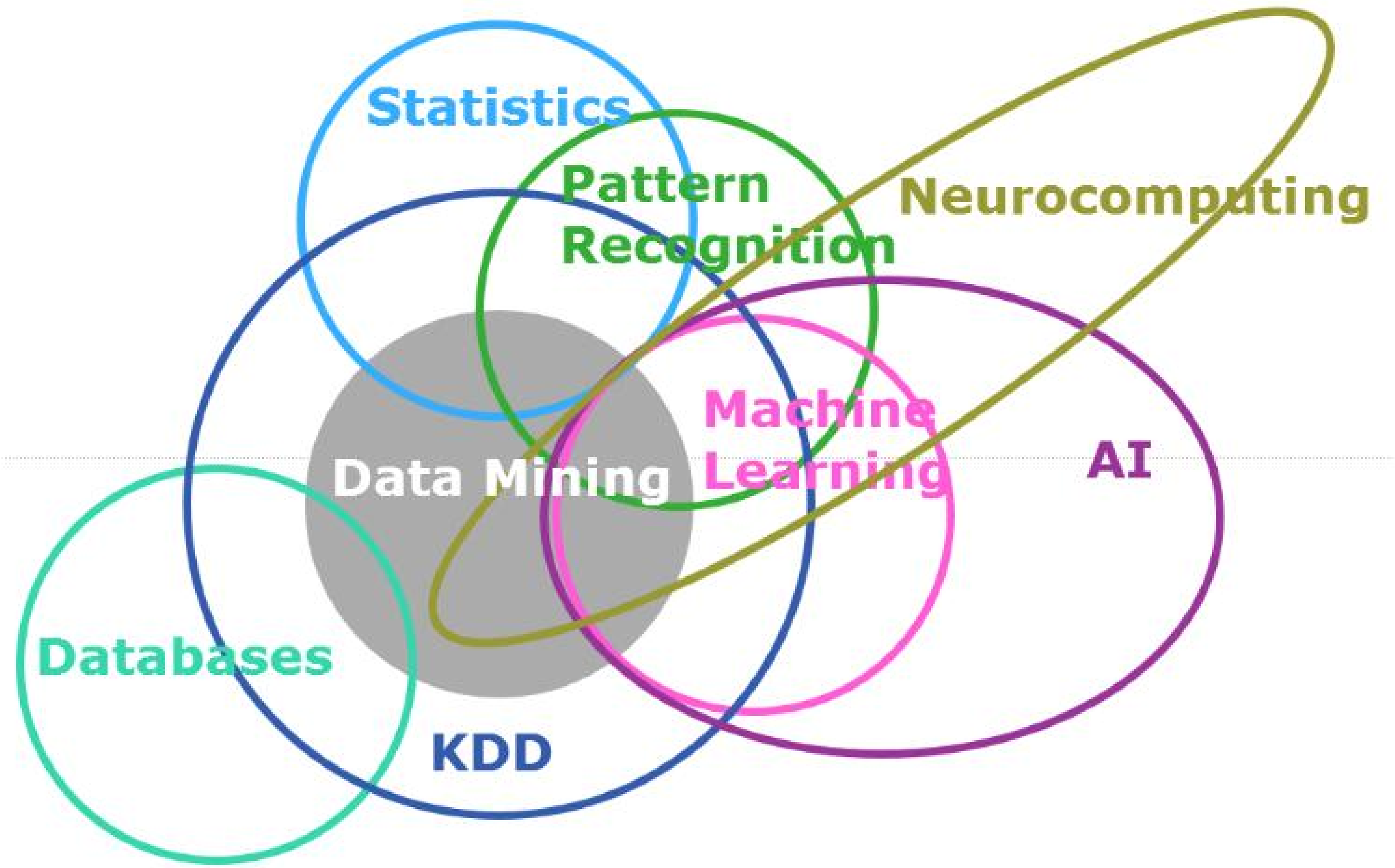
> “
> *Give machines the ability to learn without explicitly programming them*
>
> - Arthur Samuel, 1955

Arthur Samuel is recognized for developing the first learning machine, which was capable of
playing and winning checkers.

His algorithms used a heuristic search memory to *learn from experience*.

By the mid 1970's his program was beating capable human players.

Data-related Disciplines



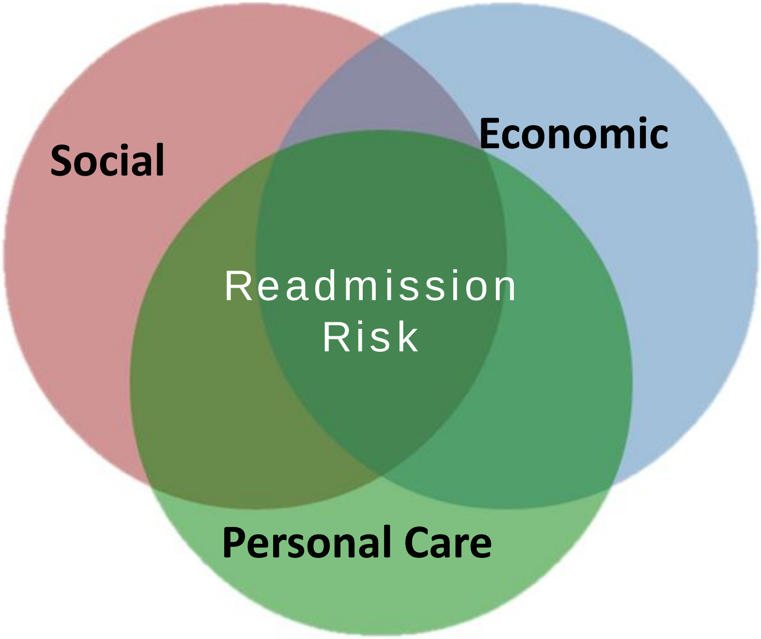
Source: SAS Institute

Predictors and Outcomes

- **Learning is the key**

- Humans naturally see patterns in data like risk factors

- ***Machine learning*** allows computers to identify those factors & more complex relationships or patterns that change over time

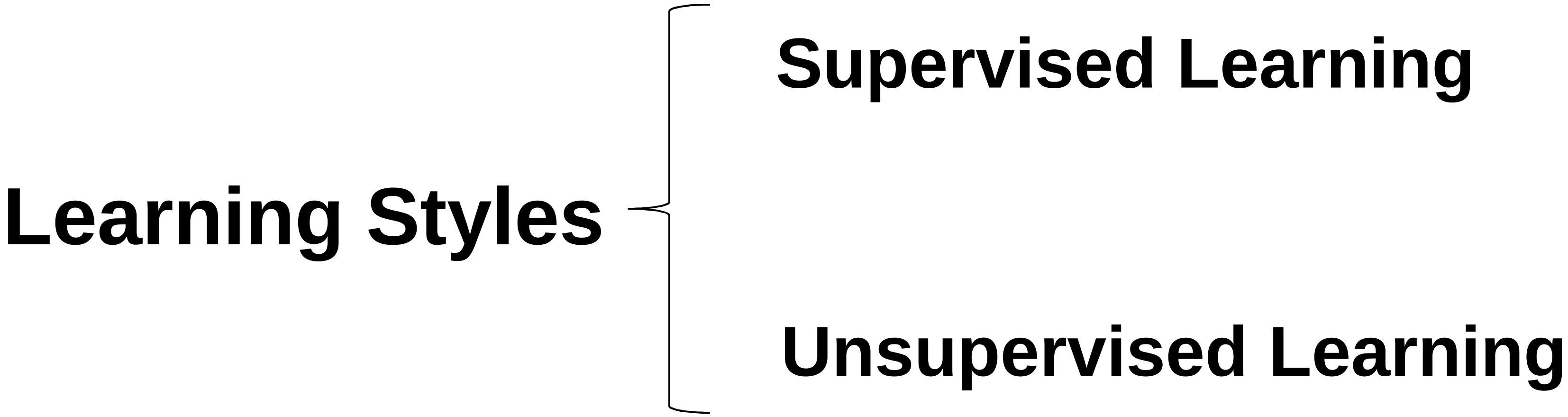


Machine Learning Impact

- **Prognosis**
  - Data can be drawn directly from EHRs or claims databases, allowing models to use thousands of rich predictor variables
  - Prediction of infections, hospital readmissions, healthcare utilization, mortality, etc
- **Interpretive Medical Specialties**
  - Interpreting digitized images, which can easily be fed directly to algorithms
  - Monitor and interpret streaming physiological data, replacing aspects of anesthesiology and critical care
- **Diagnostics**
  - Alarming frequency of diagnostic errors and the lack of interventions to reduce them
  - Suggest high-value tests, and reduce overuse of testing

Source: Obermeyer and Emanuel, NEJM 2016

Machine Learning – “Learning Styles”



Learning Styles

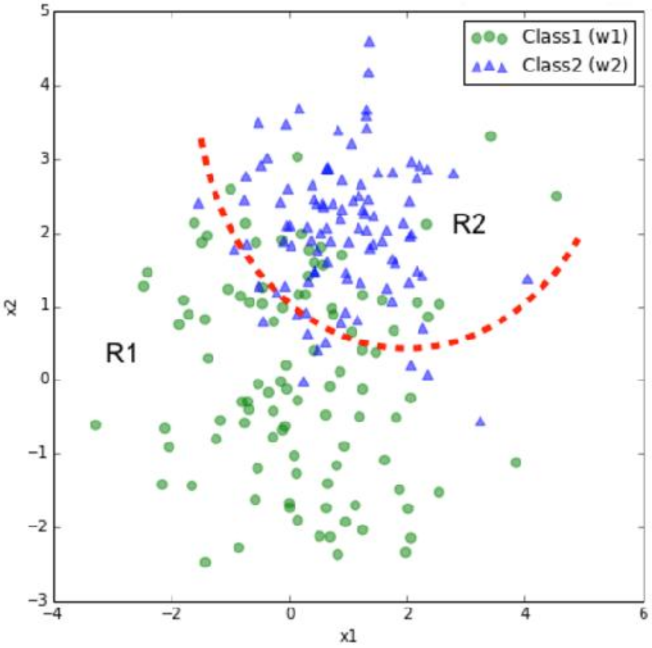
Supervised Learning

Unsupervised Learning

Supervised Learning

- The set of (training/learning) data consists of a set of input data and correct responses (labels) corresponding to every piece of data.

- The algorithm has to generalize such that it is able to correctly (or with low error margin) respond to all possible inputs.

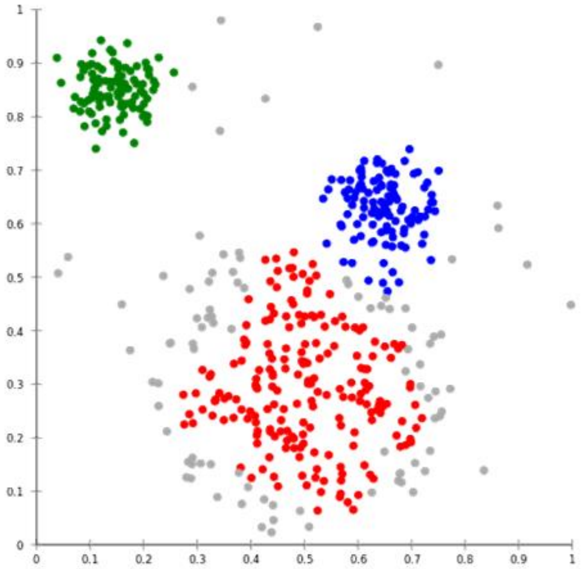


Supervised Learning Algorithms

- K-Nearest Neighbors
- Decision Trees
- Linear Regression
- Logistic Regression
- Random Forests
- Support Vector Machines
- Neural Networks

Unsupervised Learning

- No information about correct outputs (labels) is available
- Algorithm must determine the data patterns on its own.
- Tends to restructure the data into something else, such as new features that may represent a class or a new series of uncorrelated values. Useful in providing insights into the meaning of data and new useful inputs to supervised machine learning algorithms.

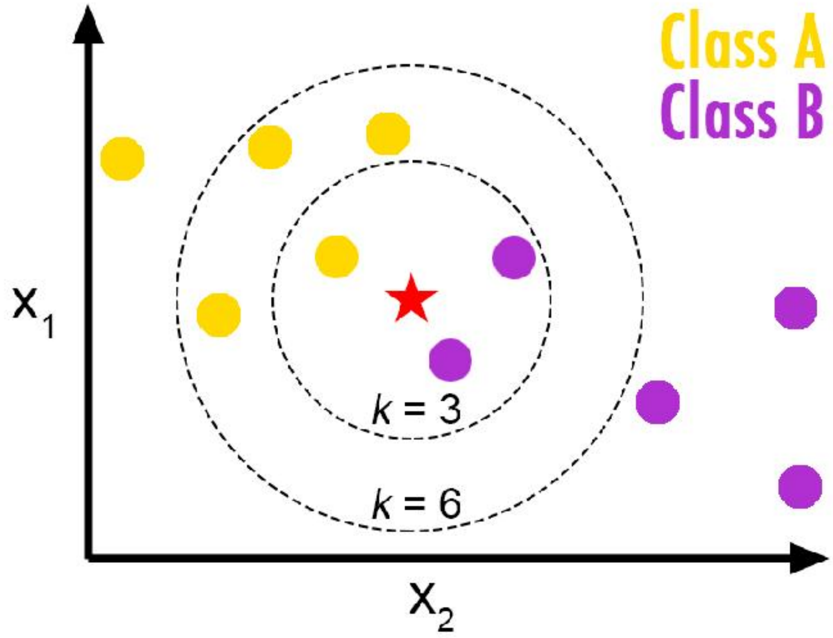


Unsupervised Learning Algorithms

- **Clustering Algorithms**
- **Principle Component Analysis**
- **Singular Value Decomposition**
- **Independent Component Analysis**

K-Nearest Neighbors – Supervised

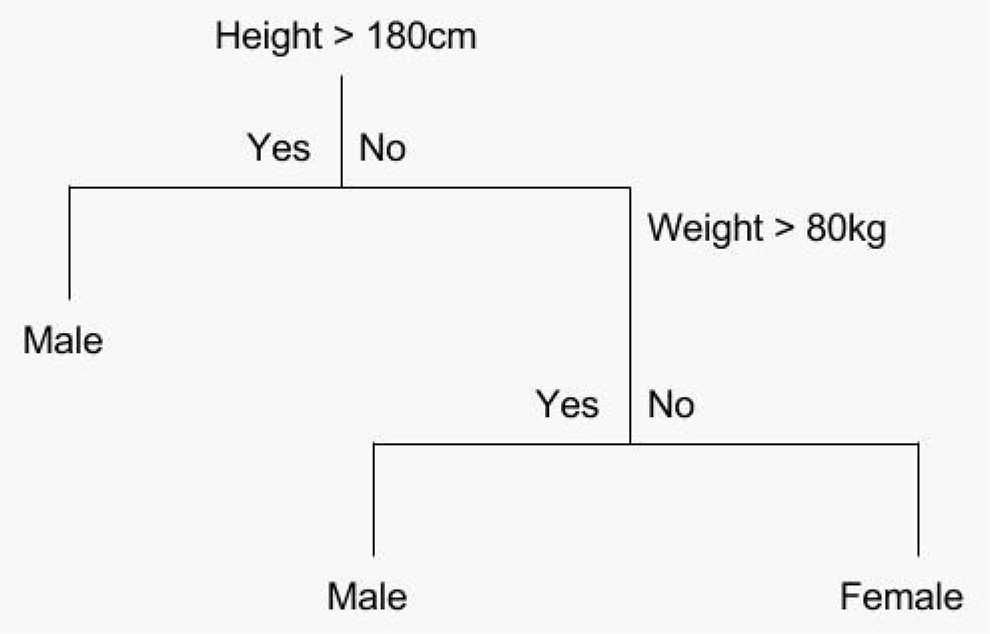
- Group cases together based on the class of the *k*-nearest neighbors, where *k* is the number of neighbors to consider
- Uses similarity matching to classify new cases based on similarity to known cases



Simple and intuitive method, but choosing optimal *k* is tricky; need large data set for accuracy, and doesn't work well when class distribution is skewed

Decision Tree – Supervised

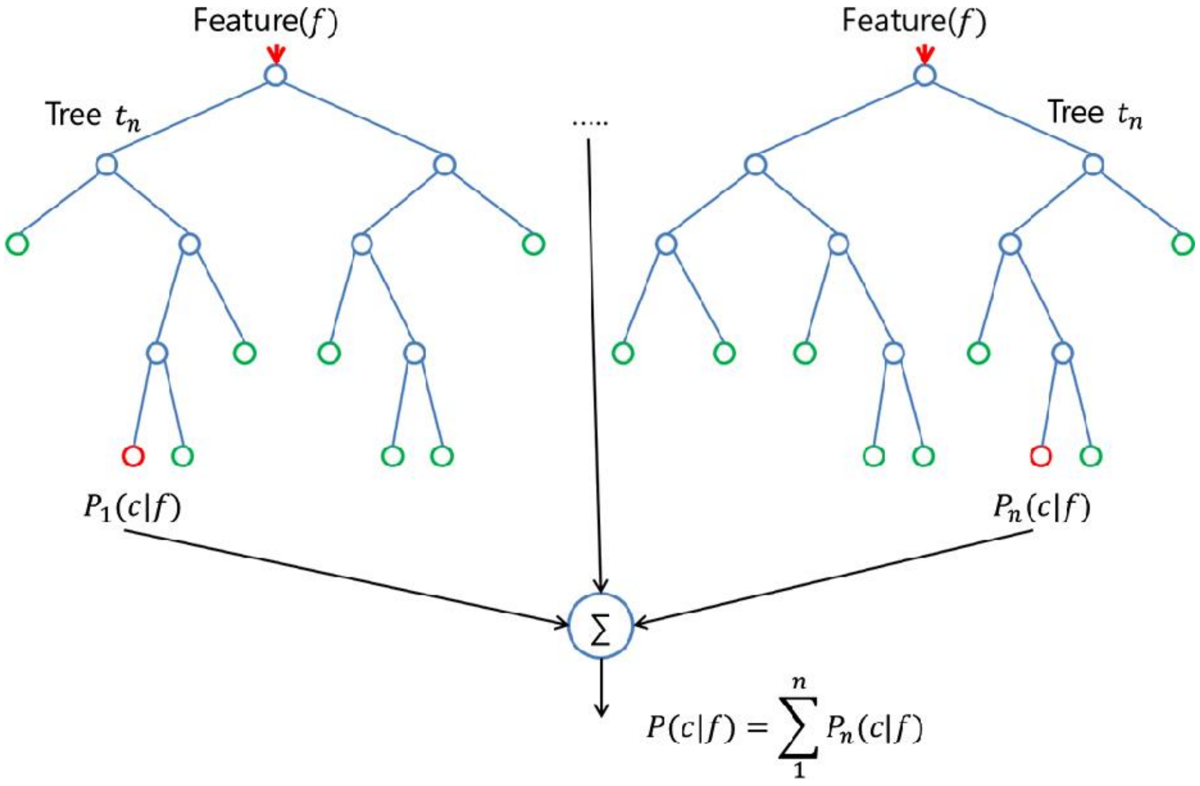
- Also referred to as Classification and Regression Trees (CART)
- Each root node represents a single input variable (x) and a split point on that variable (assuming the variable is numeric). The leaf nodes of the tree contain an output variable (y) which is used to make a prediction.



Various approaches may be used in the selection of which input variable to use and the specific split or cut-point.

Random Forests – Supervised

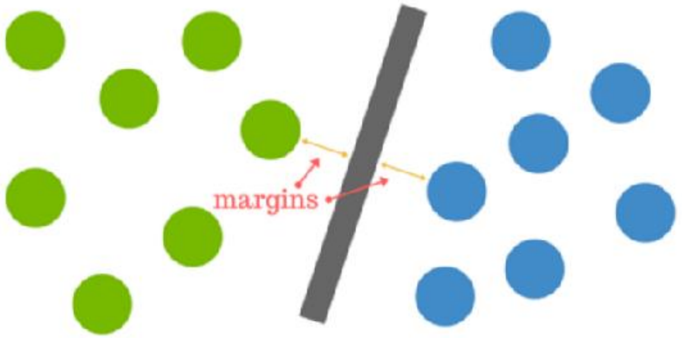
- Resolves the overfitting issue for single decision trees.
- Training of large number of trees (>500) using random sample of training set and random subset of features, and then classifies a new case based on “majority rules”.

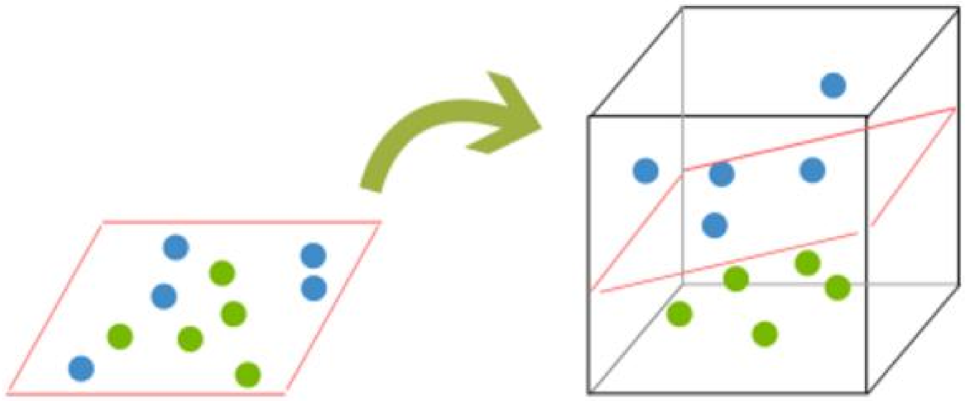


A single decision tree can be easily understood, however, with a large number of trees, difficult to trace decision logic

Support Vector Machines – Supervised

- SVMs are based on the idea of finding a hyperplane that best divides a dataset into classes.
- Choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set.

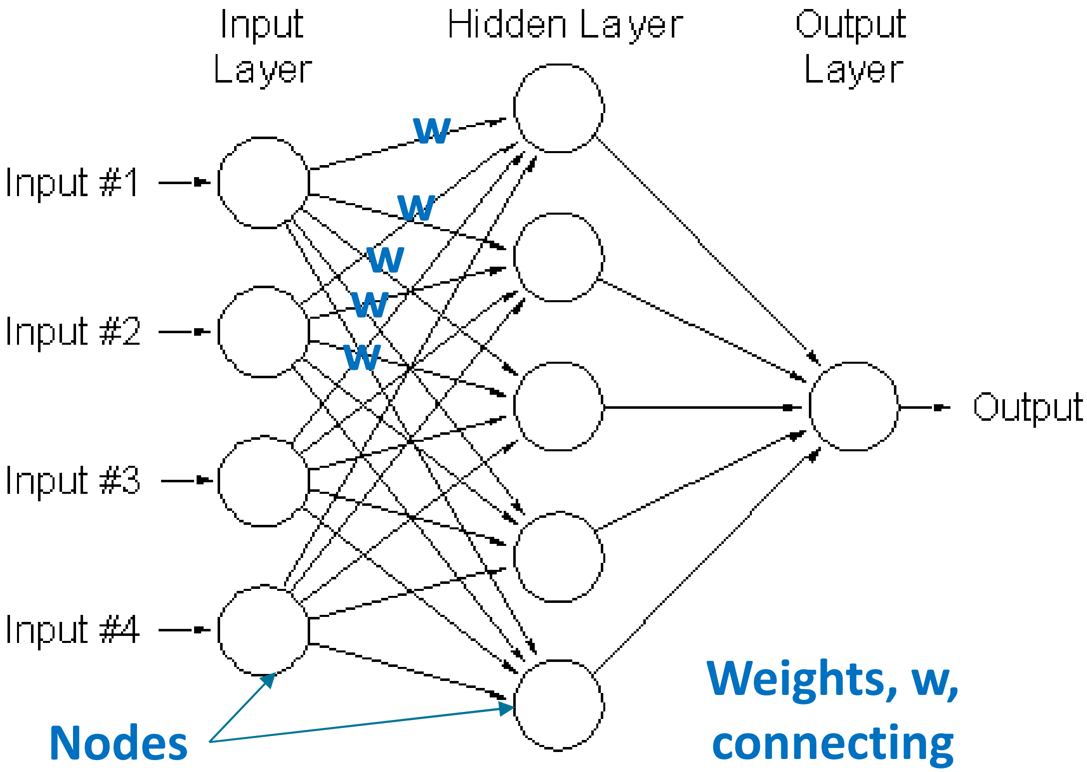




Data will continue to be mapped into higher and higher dimensions through kernel transformations until a hyperplane can be formed to segregate it

Neural Networks - Supervised

- Inspired by the structure and functional aspects of neurons in the brain.
- Weights that strengthen or weaken a connection link each node to the node in the next layer.
- Errors can be propagated back from the output to the input to adjust weights.



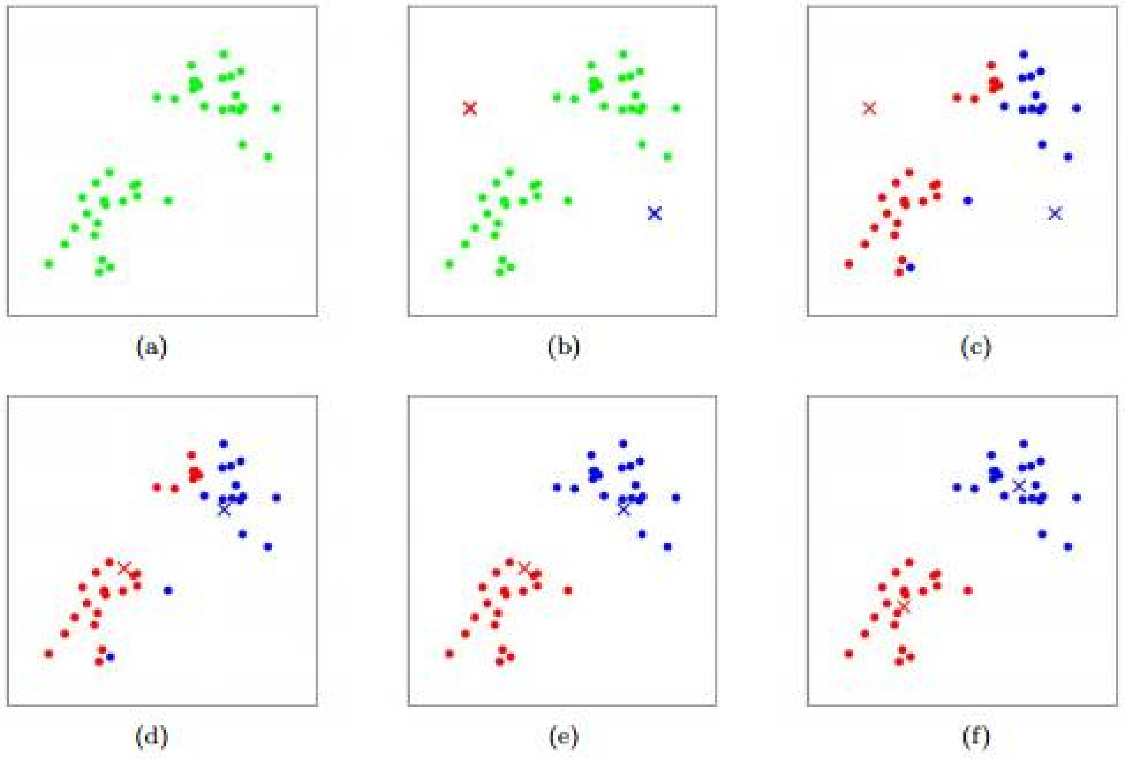
Nodes

Weights, w, connecting each node

“Black box” model – difficult to explain decision logic, however, with increasingly complex machine learning tools available, this issue is no longer uniquely tied to neural networks

K-Means Clustering – Unsupervised

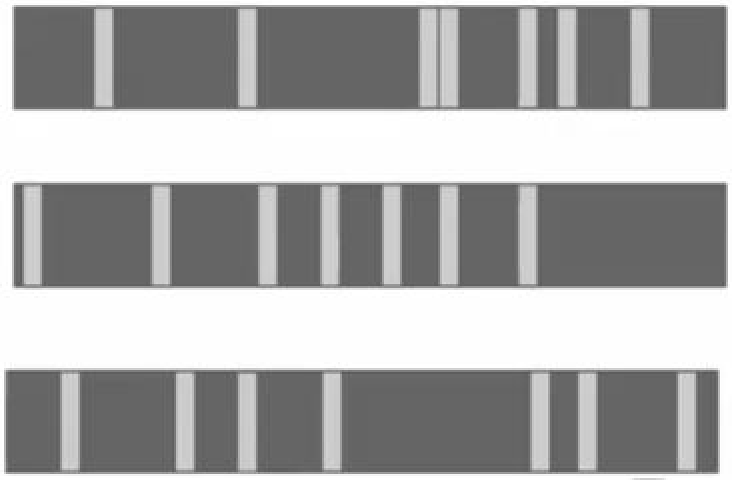
- Randomly select *k* observations from the entire data set as preliminary cluster centroids, where *k* is the number of clusters.
- Assign remaining cases into *k* clusters by minimizing a within-cluster distance function, then recalculate cluster's centroid, and REPEAT until convergence .
- Can be used as a data pre-processing step for other algorithms or to create new features.



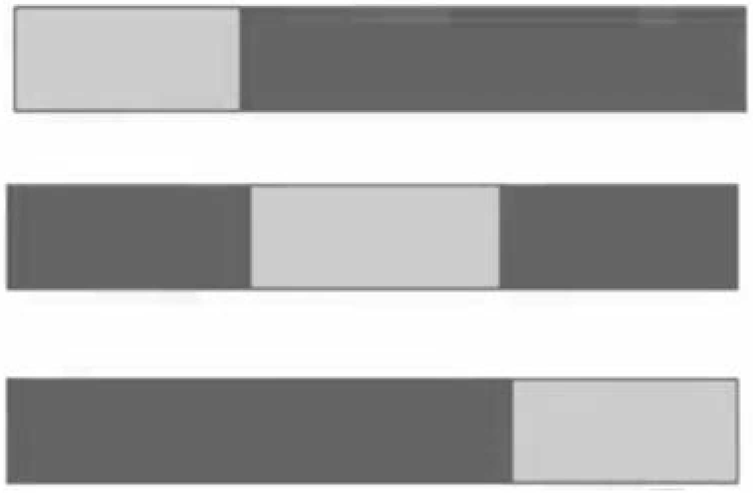
Based on assumption that clusters are spherical, separable and similarly sized, which may not be true for the data

Cross-Validation

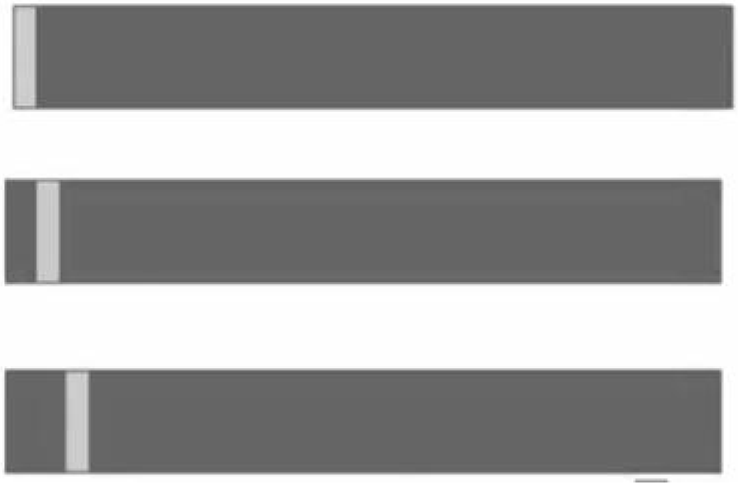
Random Subsampling



K-fold



Leave one out



Receiver Operating Characteristics (ROC)

Confusion Matrix

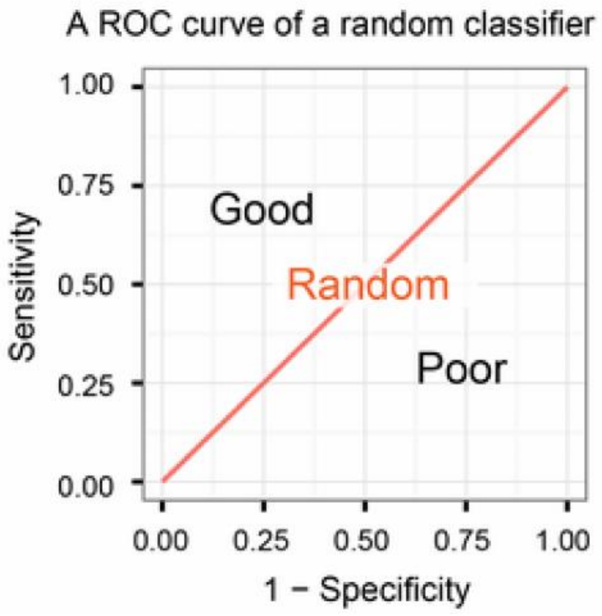
		True class	
		p	n
Hypothesized class	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

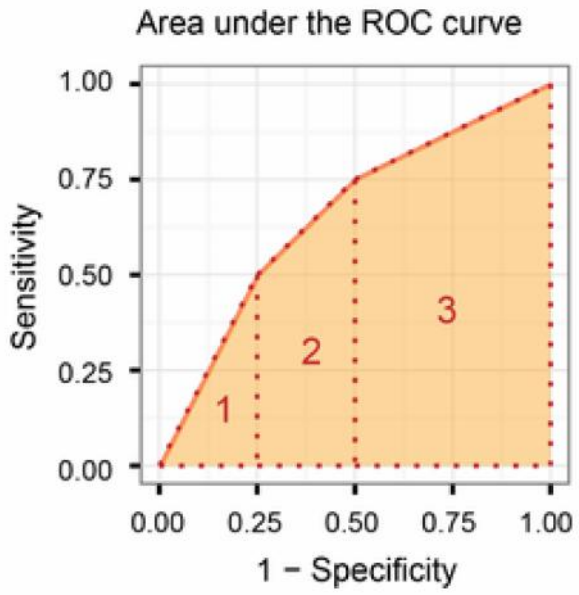
$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$


A ROC curve of a random classifier


Area under the ROC curve

Source: Fawcett,
Pattern Recog. Letters, 2006

Machine Learning in Practice

Uses

Pattern recognition	Flu season onset and duration
Classification	High or low risk of developing diabetes
Detection	Identify patients who have severe sepsis before clinician does
Prediction	Estimate a patient's risk of hospital readmission

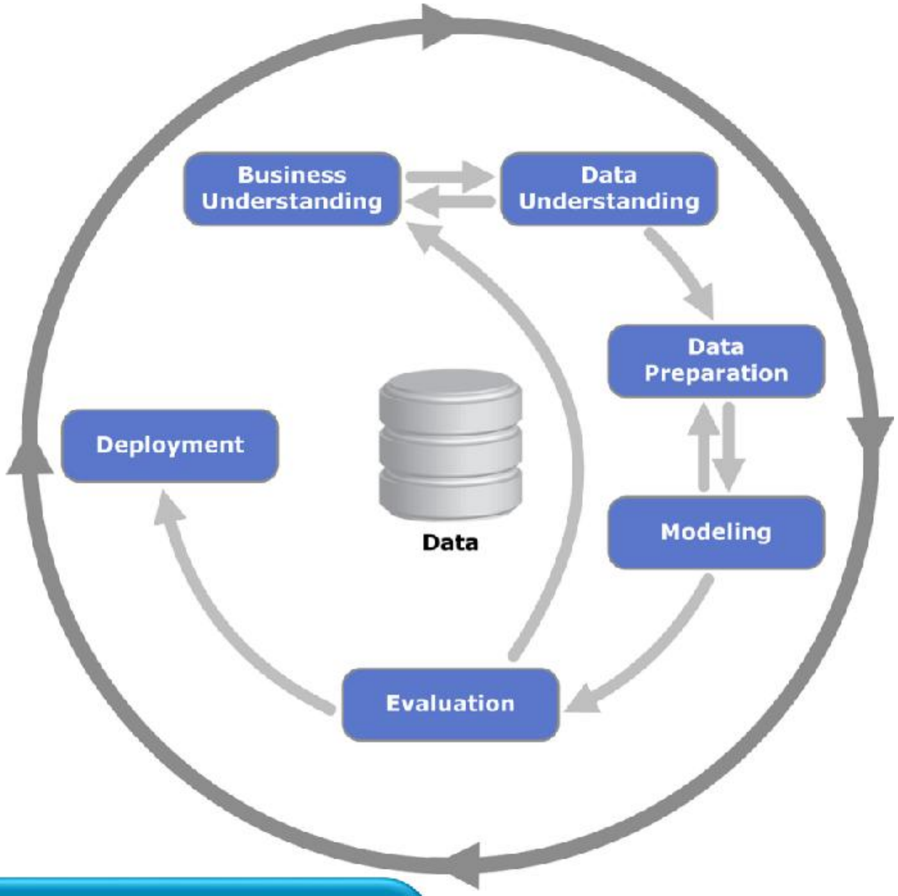
Limitations

- Availability of suitable data
- Sample data set size
- Underlying biases in data sets that are not representative of the target population
- Overfitting model to training data
- "Needle in a haystack" results in undesirable false positives

Big data has not solved all of the data-related issues

Model Development Workflow

- Identify the problem
- Explore data and identify potential features of interest
- Select the model type based on the data available and the model's goal
- Apply model method to the data
- Assess classification performance using cross-validation
- Deploy model in test environment
- After confirming that model performs as expected, roll out to production environment



Frequent interactions with domain experts & target users throughout the process is necessary to achieve success

Prediction and Causality

HERMAN/Unger
9/5
Unger

Rapid pulse, sweating, shallow breathing. According to the computer, you've got gallstones

Who can do Machine Learning?

- Computer Scientists
- Engineers
- Physicists
- Mathematicians
- Data Science Enthusiasts
- Others

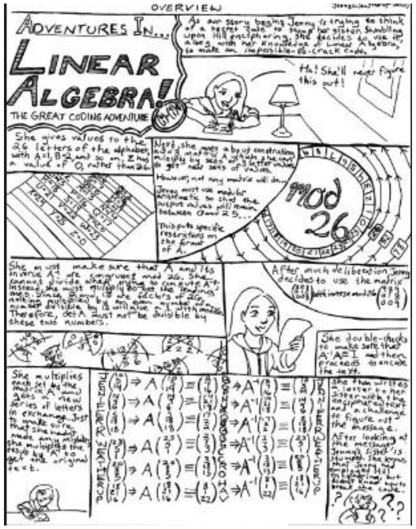
Pre-requisites for Machine Learning

Programming

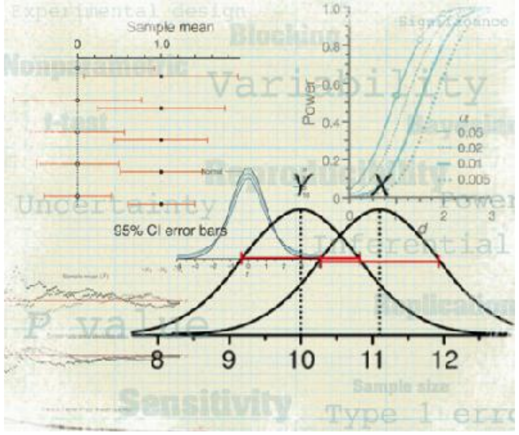
```
i a.length&&(x=a[i])&&x.oSrc;i
d.MM_p) d.MM_p=new Array();
preloadImages.arguments; for(i=0; i a.length;
d.MM_p[j]=new Image; d.MM_p[j++].src=a[i];
if((p=n.indexOf("?"))>0&&parent.frames.length
document; n=n.substring(0,p);}
for (i=0;!x&&i<d.forms.length;i++)
d.layers.length;i++) x=MM_findObj(n,d.layers[i].document
d.getElementById(n); return x;}
document.MM_sr=new Array; for(i
document.MM_sr[j++]=x; if(!x.oSrc)
```

Math

(Calculus, Linear Algebra)



Statistics



Fundamentals

1. Planning and Data Collection

2. Data Assumptions and Preprocessing

3. Interpreting model results

4. Improving and tuning models

5. Driving to business value

Tools

Python: Scikit-Learn

R: Caret

Public Datasets

UCI Machine Learning Repo

Kaggle

Data.gov

Data Governance

- Greater amounts of data are not helpful if they are not of **good quality and easily exchangeable**. Good quality data are **accurate, complete, timely, relevant, and consistent**; **adherence to standards** for data quality would ensure the reliability of both data and analytics.

Why Data Governance?

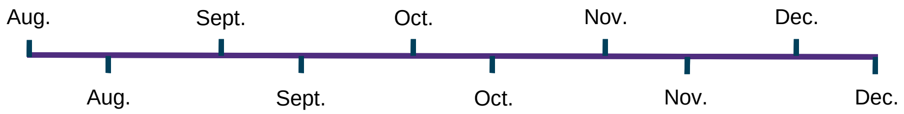
Battle of Austerlitz, December 2, 1805

Napoleon vs. The Third Coalition (Britain, Russia, Austria)

The plan was for the Russian and Austrian armies to meet on Oct. 20.

However, the Allied planners failed to realize that the Russians were using the Julian calendar and the Austrians were using the Gregorian calendar.

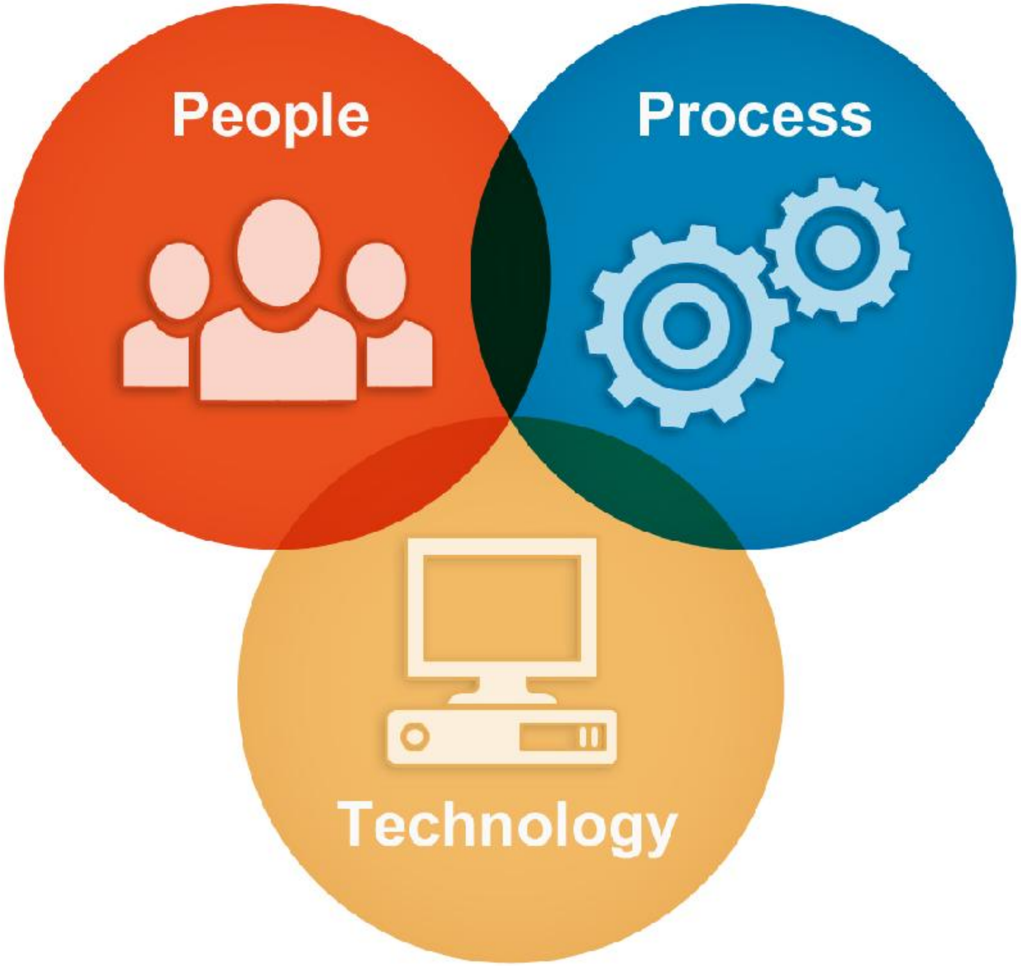
The calendars were 12 days apart.



The Austrians were defeated without Russian support; the Russians arrived 'late' and were defeated by Napoleon.

Common terminology could have turned the tide in this battle!

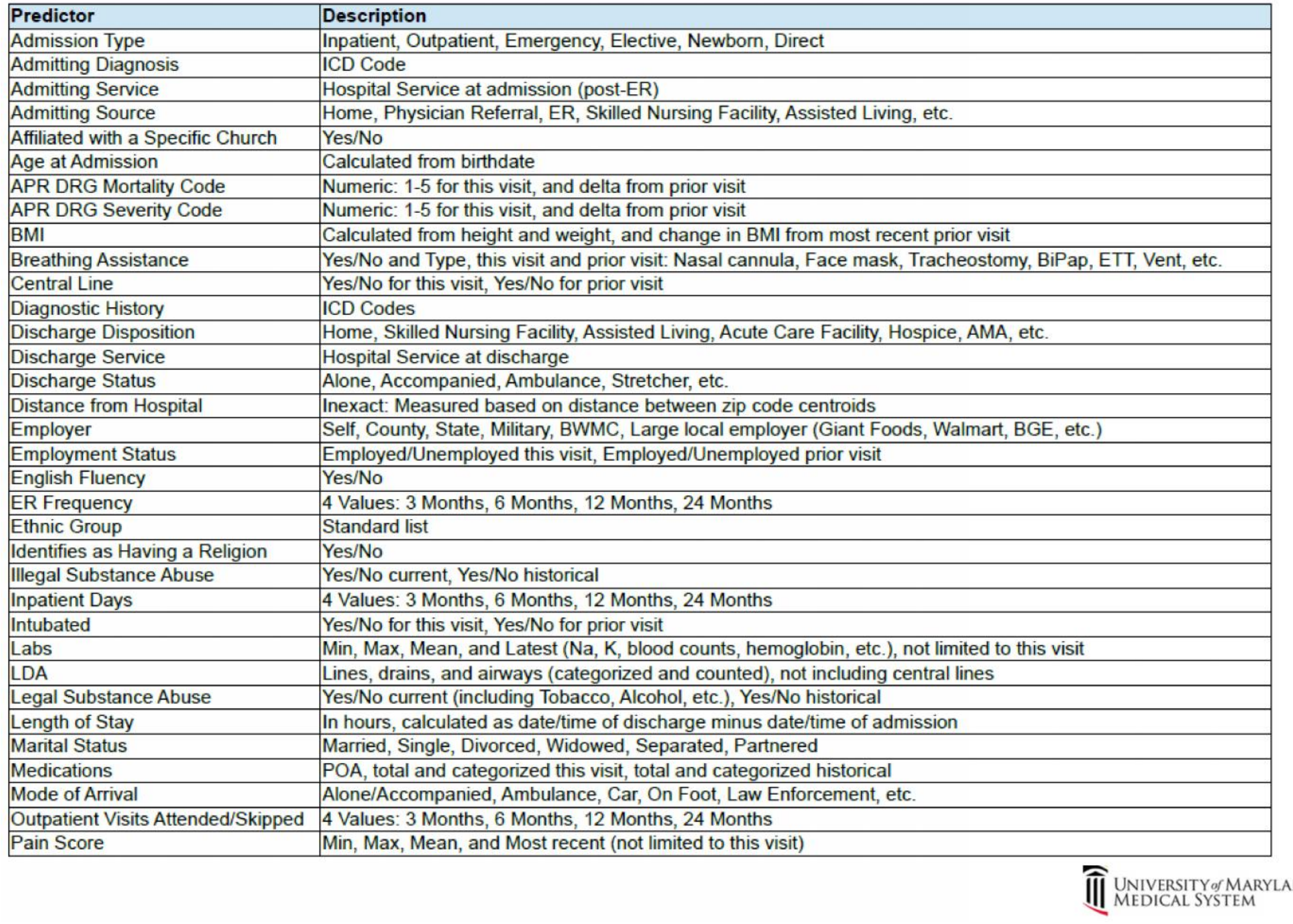
Data Governance Components



30-Day Hospital Patient Risk Stratification

- *Using clinical and non-clinical variables, stratify patients by their risk of readmission to the hospital within 30 days of discharge.*
- *Provide stratified list of patients in EMR for care management programs.*

Input Variables



Predictor	Description
Admission Type	Inpatient, Outpatient, Emergency, Elective, Newborn, Direct
Admitting Diagnosis	ICD Code
Admitting Service	Hospital Service at admission (post-ER)
Admitting Source	Home, Physician Referral, ER, Skilled Nursing Facility, Assisted Living, etc.
Affiliated with a Specific Church	Yes/No
Age at Admission	Calculated from birthdate
APR DRG Mortality Code	Numeric: 1-5 for this visit, and delta from prior visit
APR DRG Severity Code	Numeric: 1-5 for this visit, and delta from prior visit
BMI	Calculated from height and weight, and change in BMI from most recent prior visit
Breathing Assistance	Yes/No and Type, this visit and prior visit: Nasal cannula, Face mask, Tracheostomy, BiPap, ETT, Vent, etc.
Central Line	Yes/No for this visit, Yes/No for prior visit
Diagnostic History	ICD Codes
Discharge Disposition	Home, Skilled Nursing Facility, Assisted Living, Acute Care Facility, Hospice, AMA, etc.
Discharge Service	Hospital Service at discharge
Discharge Status	Alone, Accompanied, Ambulance, Stretcher, etc.
Distance from Hospital	Inexact: Measured based on distance between zip code centroids
Employer	Self, County, State, Military, BWMC, Large local employer (Giant Foods, Walmart, BGE, etc.)
Employment Status	Employed/Unemployed this visit, Employed/Unemployed prior visit
English Fluency	Yes/No
ER Frequency	4 Values: 3 Months, 6 Months, 12 Months, 24 Months
Ethnic Group	Standard list
Identifies as Having a Religion	Yes/No
Illegal Substance Abuse	Yes/No current, Yes/No historical
Inpatient Days	4 Values: 3 Months, 6 Months, 12 Months, 24 Months
Intubated	Yes/No for this visit, Yes/No for prior visit
Labs	Min, Max, Mean, and Latest (Na, K, blood counts, hemoglobin, etc.), not limited to this visit
LDA	Lines, drains, and airways (categorized and counted), not including central lines
Legal Substance Abuse	Yes/No current (including Tobacco, Alcohol, etc.), Yes/No historical
Length of Stay	In hours, calculated as date/time of discharge minus date/time of admission
Marital Status	Married, Single, Divorced, Widowed, Separated, Partnered
Medications	POA, total and categorized this visit, total and categorized historical
Mode of Arrival	Alone/Accompanied, Ambulance, Car, On Foot, Law Enforcement, etc.
Outpatient Visits Attended/Skipped	4 Values: 3 Months, 6 Months, 12 Months, 24 Months
Pain Score	Min, Max, Mean, and Most recent (not limited to this visit)

Data Processing

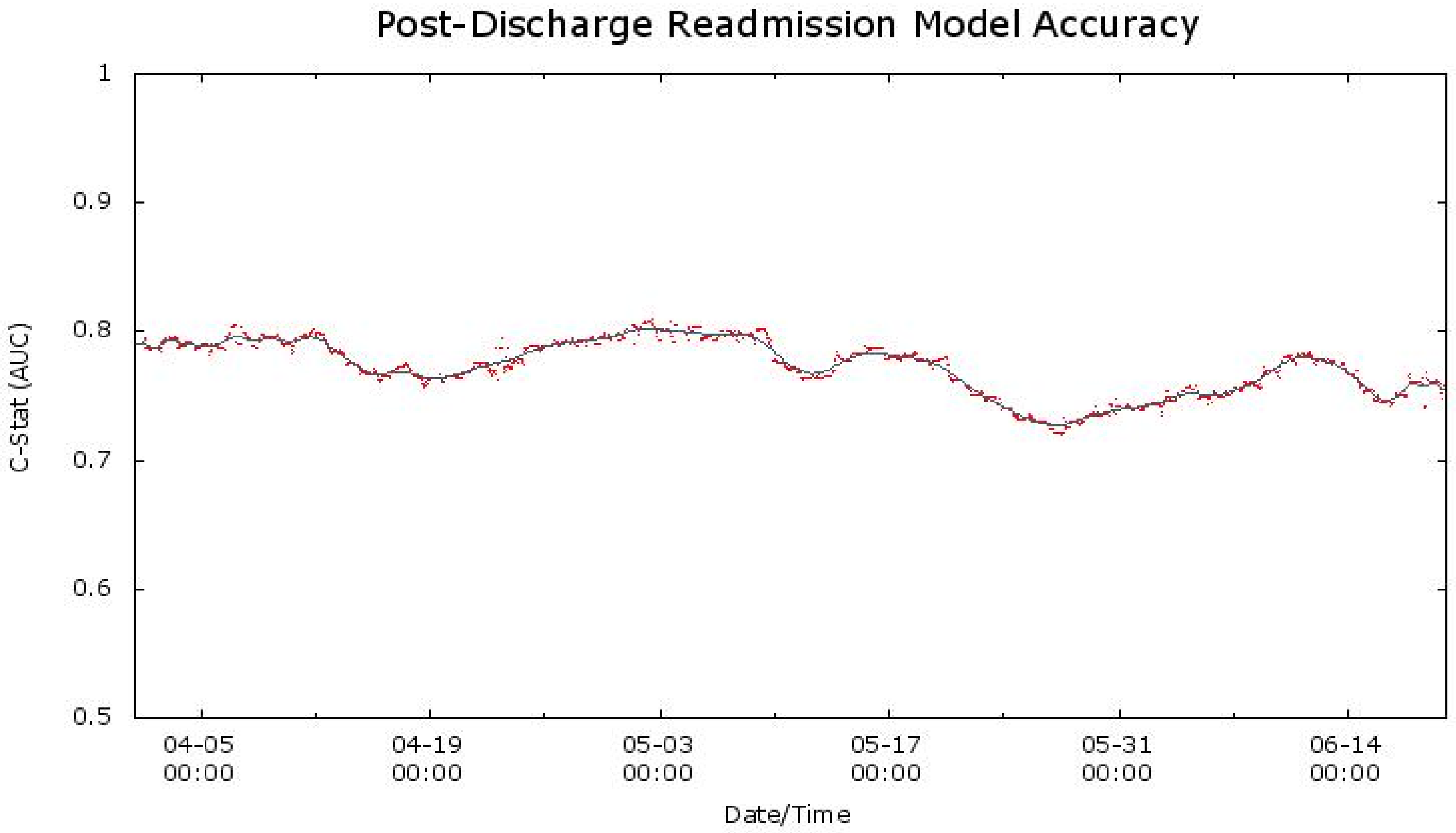
- Started with 8800 variables
- Significant data cleaning was required
- Final data set has 382 features
- Used 12-way cross-validation

  with 80/20 data splits to validate all models

Models Attempted

- K-Nearest-Neighbor (Knn)
- Classification and Regression Trees
- Recursive Partitioning (Random Forest)
- Bayes Additive Regression Trees (BART)
- Gradient Boosted Regression Trees (GBRT)
- Matrix Factorization
- Support Vector Regression (SVR - using both linear and RBF kernels)
- Linear Regression
- Logistic Regression
- Classic Neural Network
- Convolutional Neural Network (CNN – sometimes called “Deep Learning”)

Model Results



UNIVERSITY *of* MARYLAND MEDICAL SYSTEM

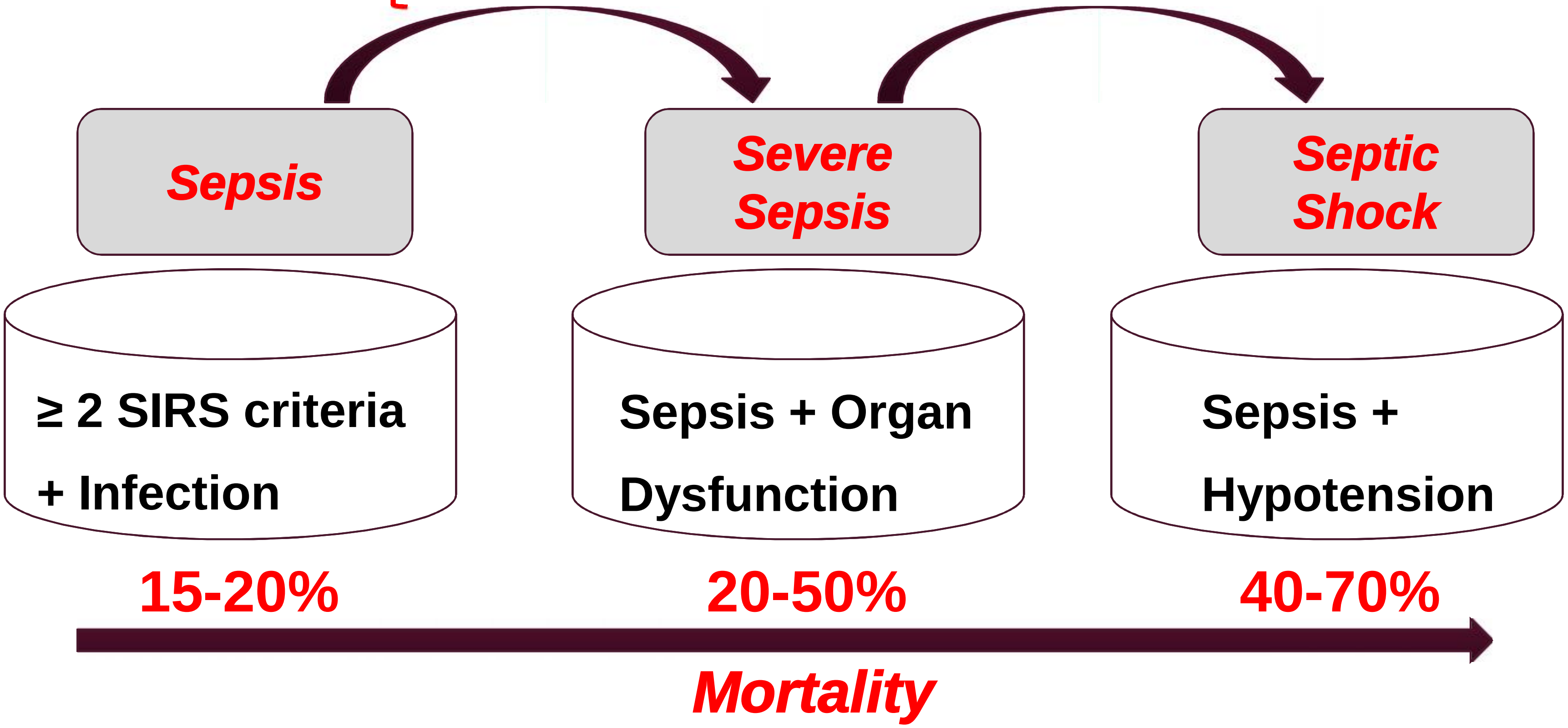
Early Identification of Severe Sepsis and Septic Shock

- *Using clinical and laboratory data in real time, identify patients at risk of progressing to severe sepsis and septic shock, prior to actual onset.*
- *Facilitate EMR-based provider alerts*

Sepsis Categories

SIRS Criteria

(1) Heart rate >90 beats/min,
(2) Respiratory rate (RR) >20breaths/min (or SIRS criteria partial pressure of arterial CO_2 < 32)
(3) Temp >38°C or <36 °C
(4) WBC either >12,000 or <4,000 cells/mm^3



Sepsis

≥ 2 SIRS criteria + Infection

15-20%

Severe Sepsis

Sepsis + Organ Dysfunction

20-50%

Septic Shock

Sepsis + Hypotension

40-70%

Mortality

Quantitative Results

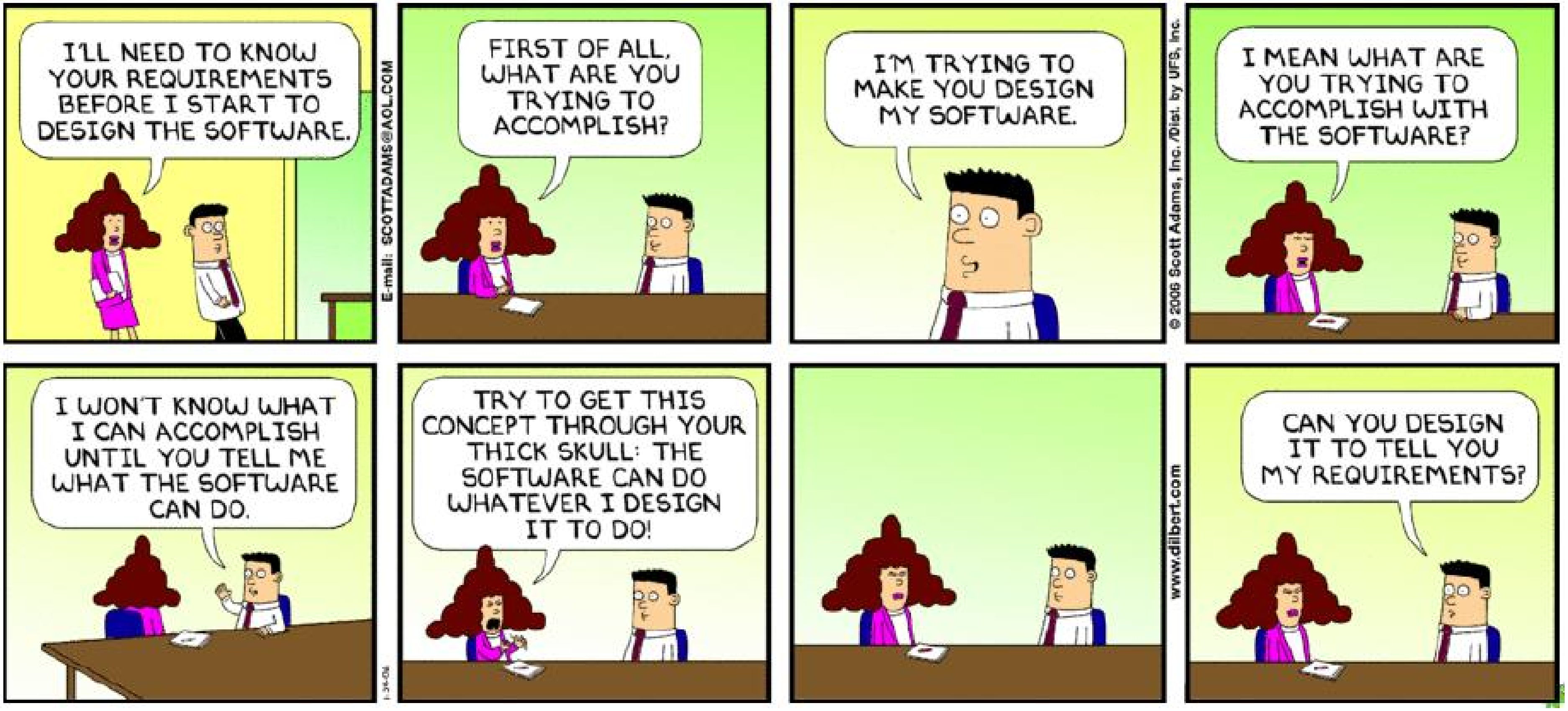
Blind Validation Set
Total Population : 19,844 ED Visits
Condition Positive: 140 (Prevalence: 0.007055)
Condition Negative: 19,704

Models Compared
LMT - Linear Model Trees
SVM - Support Vector Machine
GBRT1 - Gradient Boosted Regression Trees (Unweighted)
GBRT2 - Gradient Boosted Regression Trees (Weighted)

Raw Results	LMT	SVM	GBRT1	GBRT2
Predicted Positives	5696	2790	2208	1762
Predicted Negatives	14148	17054	17636	18082
True Positives	121	118	120	119
True Negatives	14129	17032	17616	18061
False Positives	5575	2672	2088	1643
False Negatives	19	22	20	21

UNIVERSITY of MARYLAND
MEDICAL SYSTEM

A Note of Caution


I'LL NEED TO KNOW YOUR REQUIREMENTS BEFORE I START TO DESIGN THE SOFTWARE.
E-mail: SCOTTADAMS@AOL.COM
FIRST OF ALL, WHAT ARE YOU TRYING TO ACCOMPLISH?
I'M TRYING TO MAKE YOU DESIGN MY SOFTWARE.
© 2006 Scott Adams, Inc./Dist. by UFS, Inc.
I MEAN WHAT ARE YOU TRYING TO ACCOMPLISH WITH THE SOFTWARE?
I WON'T KNOW WHAT I CAN ACCOMPLISH UNTIL YOU TELL ME WHAT THE SOFTWARE CAN DO.
TRY TO GET THIS CONCEPT THROUGH YOUR THICK SKULL: THE SOFTWARE CAN DO WHATEVER I DESIGN IT TO DO!
www.dilbert.com
CAN YOU DESIGN IT TO TELL YOU MY REQUIREMENTS?

Thank You and Questions