Introduction to Machine Learning

Mohinder Dick
Senior Software Architect
UPMC Enterprises
June 11, 2016

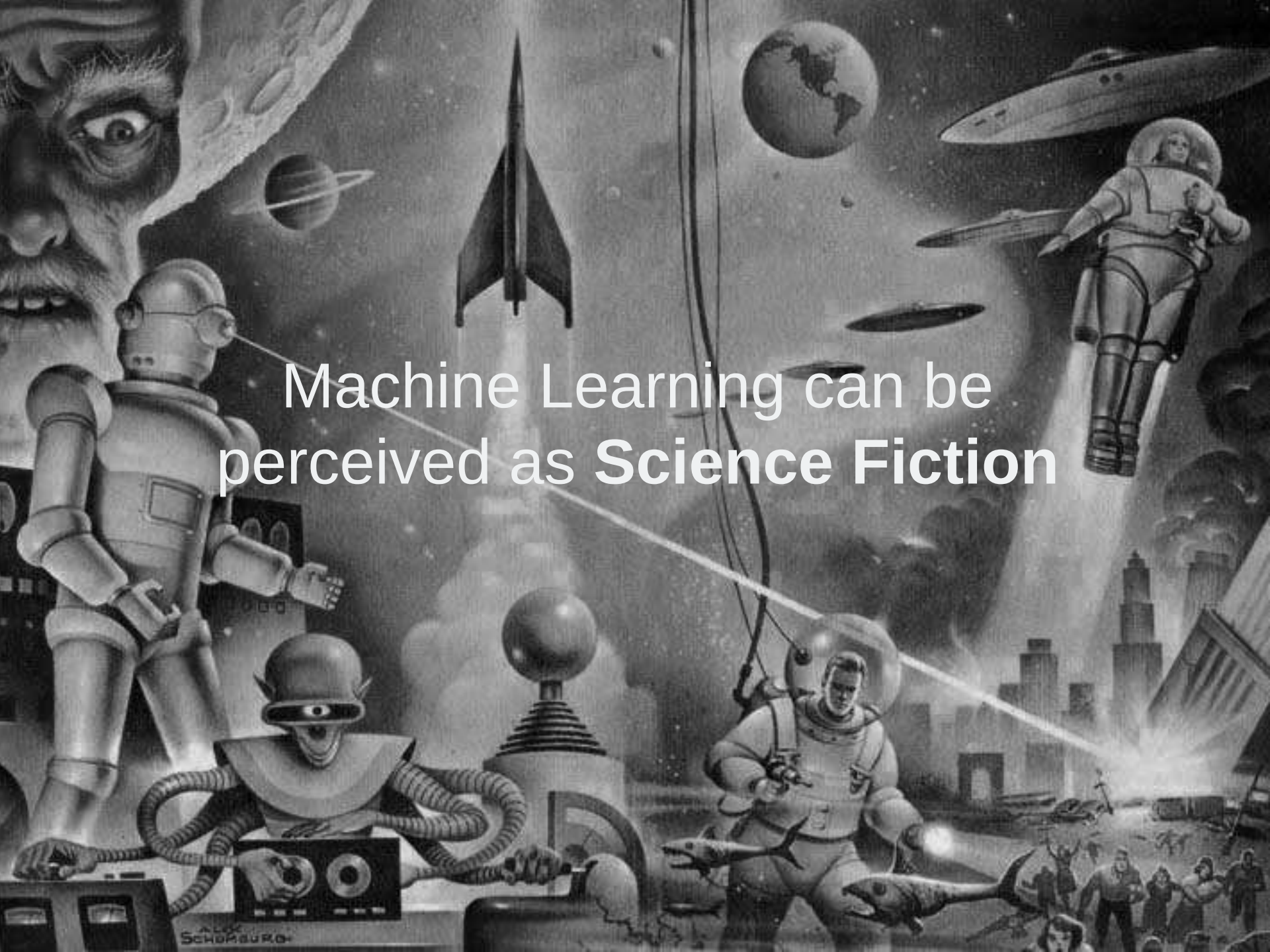
UPMC Enterprises

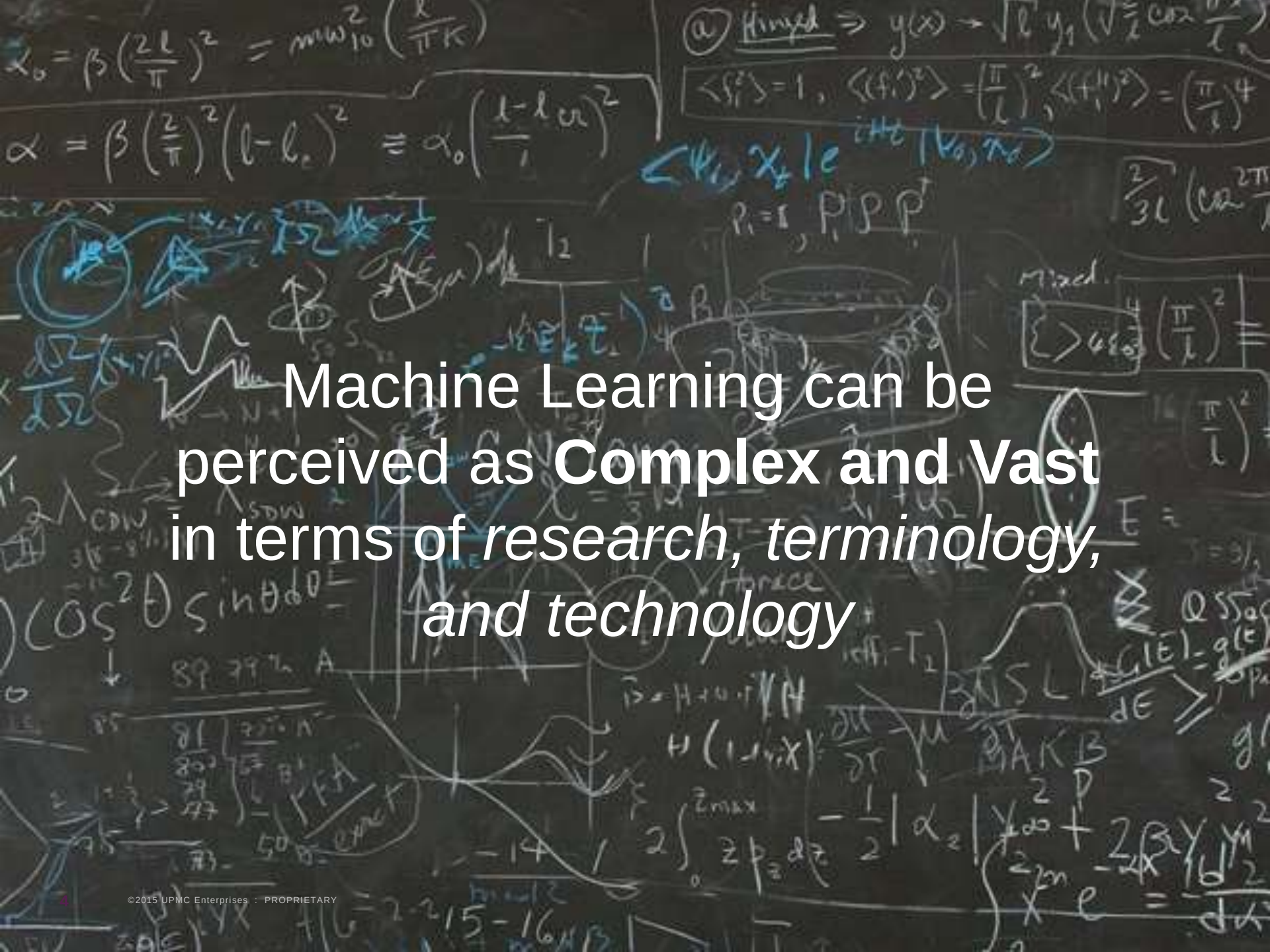
Goal

*Breakdown the perception that **machine learning** is **complex.***

Address assumptions** of what machine learning **IS**, and **what it's NOT.

Demonstrate the basics** of machine learning**.


Machine Learning can be
perceived as **Science Fiction**



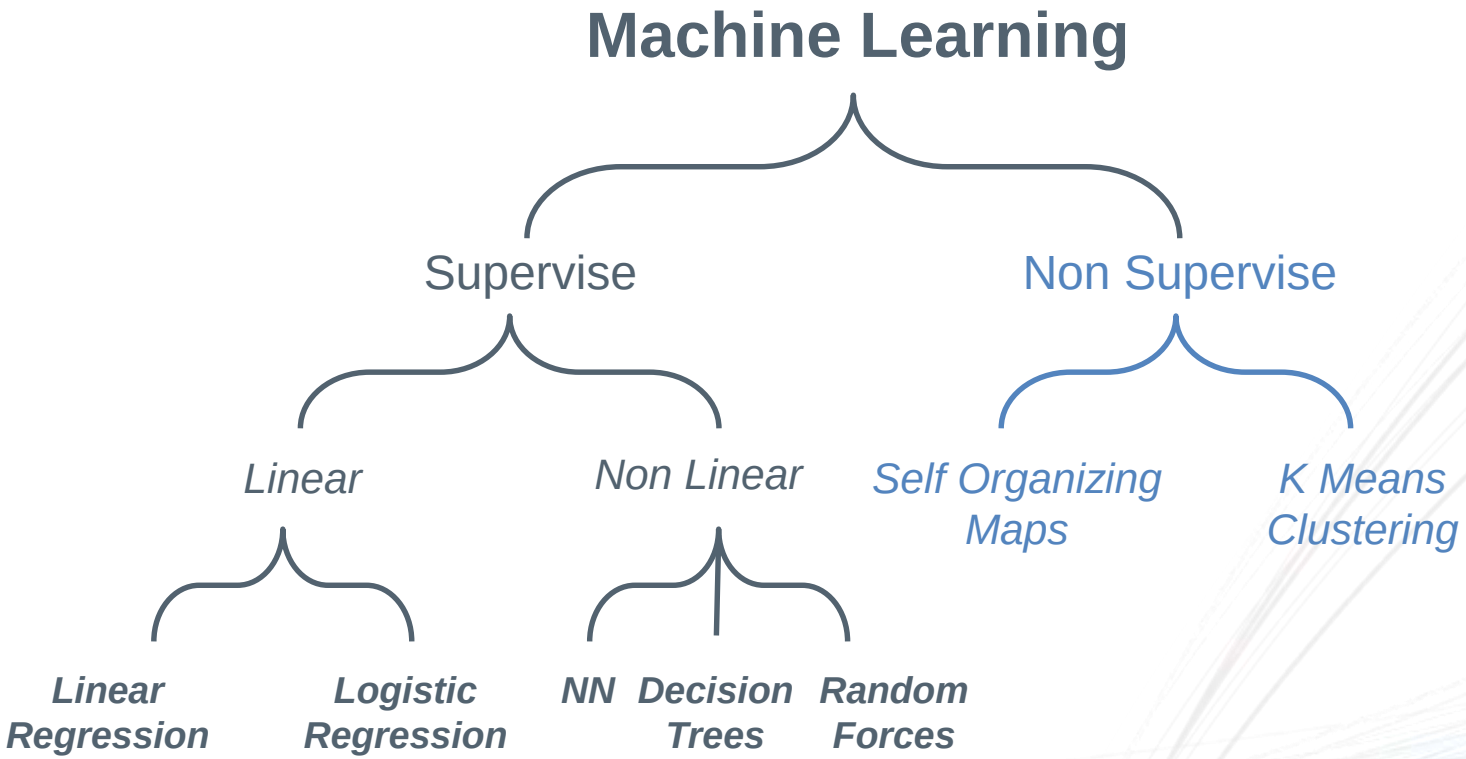
Machine Learning can be perceived as **Complex and Vast** in terms of *research, terminology, and technology*


itude: 51.5001402°
gitude: -0.126193°
Visual Analy
****
Carbon based
Humanoid...
Geek
Threat Level
TARGET ACQUIRED

What we'll accomplish today

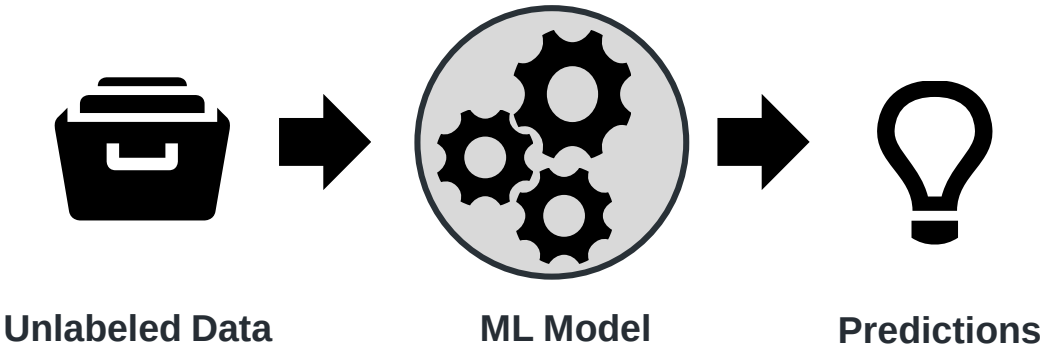
- Machine Learning Basics in 6 Slides
- How to get started
- Let's teach machines (Demo)
- Machine Learning in Healthcare
- Q & A
- References

Machine Learning in 8 Slides | **Overview**



Term	Description
Machine Learning	Algorithms that automatically improve with data
Model	Statistical representation of the algorithm's experience
Supervised Learning	Algorithms that improves using labeled examples
Linear Algorithm	Predictions are proportional to the feature/input values
Feature	Data point that affect the *target* you are trying to predict.
Target Variable	Data point that you are trying to predict
Classification	Predicting a binary or categorical target variable
Regression	Predicting a continuous target variable

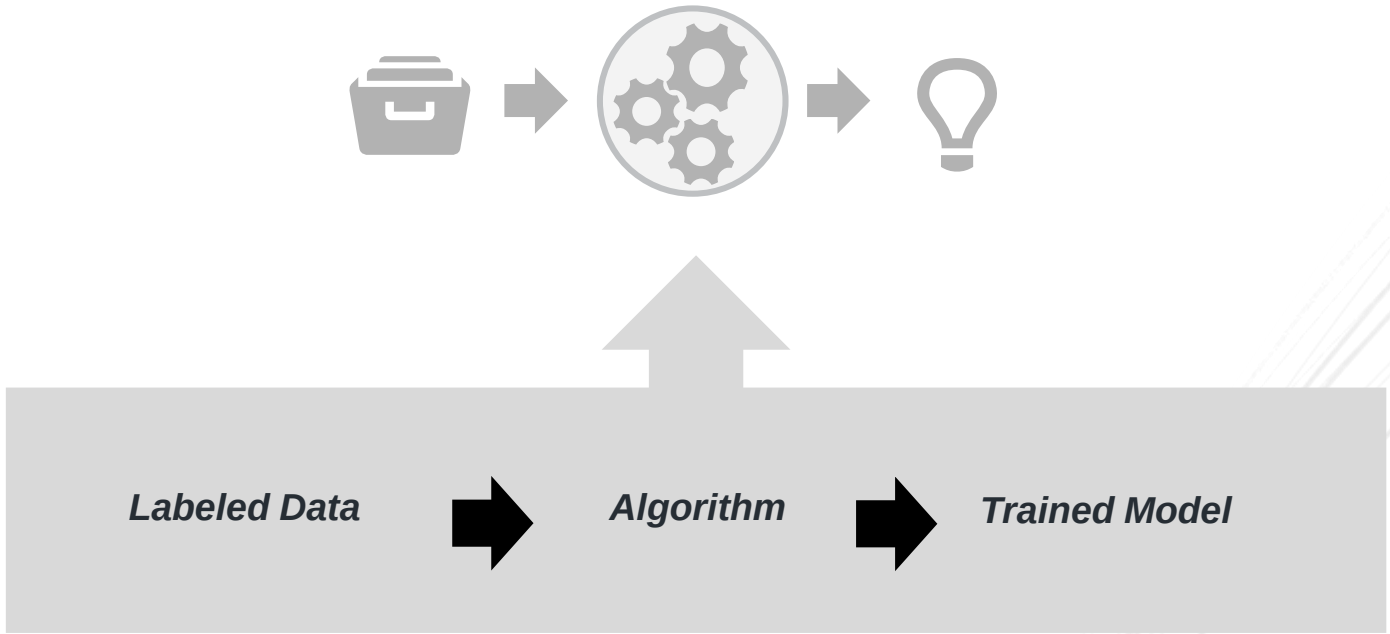
Machine Learning in 8 Slides | **Supervised Models**



- ML algorithms process “unseen” data to give predictions
- A model is a statistical representation of the algorithms experiences/training

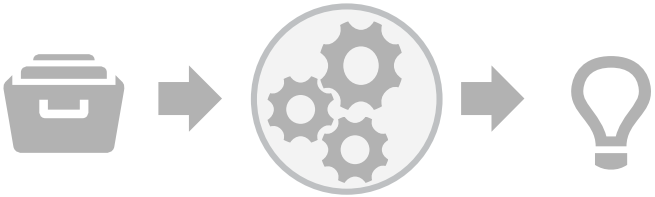
How do you get a model?

Machine Learning in 8 Slides **| Supervised Models**

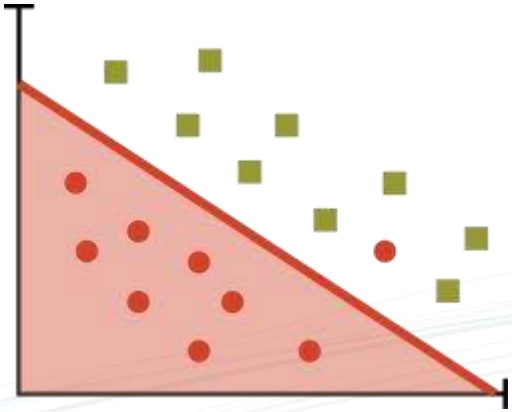


- Supervised ML algorithms get their name because they learn with help
- You have to provide them experience in labeled examples
- The algorithm translates that data into a representation called a model
- It can be used later to make predictions
- The more data the better

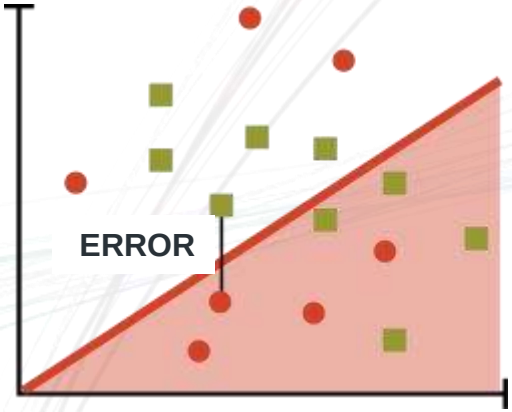
Machine Learning in 8 Slides | **Linear Models**



Labeled Data → *Algorithm* → *Trained Model*

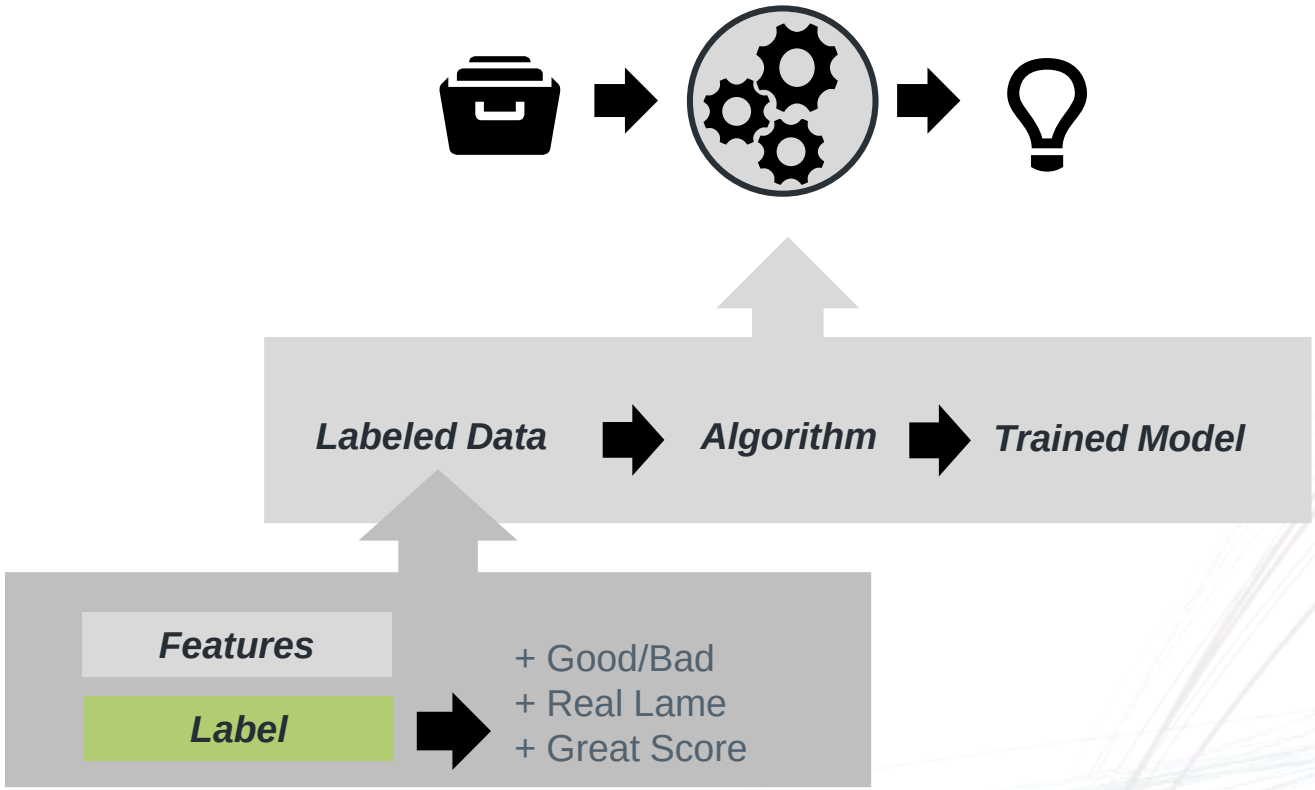


Linear Classification



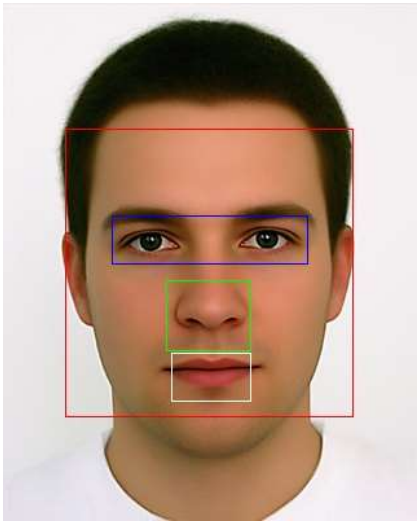
Linear Regression

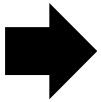
Machine Learning in 8 Slides | **Labels & Features**

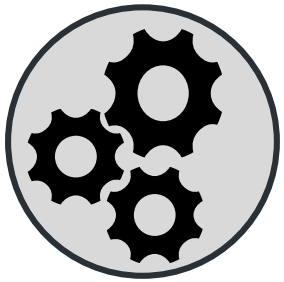


- During training features with labels are given to the algorithm to generate the model
- Both features and labels can be categorical or continuous
- The type of your target affects the flavor of algorithm you chose

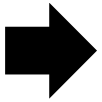
Facial Recognition





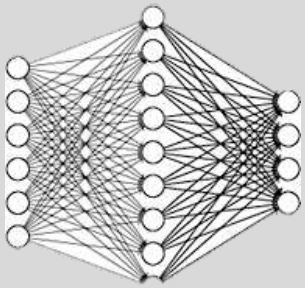


ML Model



Predictions

Geometric features called “Haars”



Non-linear model

Is this a face?

Machine Learning in 8 Slides | **Recap**

Term	Description
Machine Learning	Algorithms that automatically improve with data
Model	Statistical representation of the algorithm's experience
Supervised Learning	Algorithms that improves using labeled examples
Linear Algorithm	Predictions are proportional to the feature/input values
Feature	Data point that affect the *target* you are trying to predict.
Target Variable	Data point that you are trying to predict
Classification	Predicting a binary or categorical target variable
Regression	Predicting a continuous target variable

How to get started | **Training, Data, Tools & Approaches**

Training

Books
MOOCs
Instructor Led

Data

Government Agencies
Vendors

Tools

Data Pipeline
Spark
Python

How to get started **| Tool Details**

Apache Spark

Cluster aware execution
MLLib for machine learning

Python

Loosely typed language
Libraries – scikit, pandas and numpy

R

Statistical language, IDE ecosystem

How to get started | **Tool Details**

```
/**
 * train the model
 */
 val model = KMeans.train(train_data,CLUSTERS,ITERATIONS)

/**
 * score the data
 */
 val results = grouped_model_data.map(
   v => (
     v._1
    ,model.predict(
       Array.tabulate[Double](field_cnt)(
         i => zscore(v._2(i),sample_mean(i),sample_stddev(i))
       )
     )
   )
 )
 results.saveAsTextFile(output_path)
```

How to get started | **Data Driven Approach**

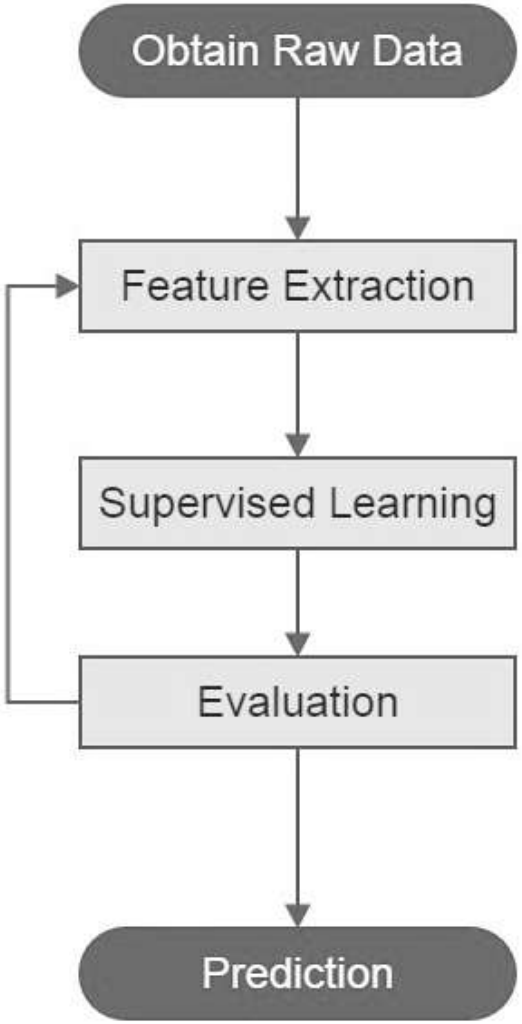
Problem
- Hospital bed capacity

Data
- Admitting Hospital
- Patient diagnosis and demographics

Analysis
- Data representation
- Data analysis
- Evaluation

Policy
- Increase capacity at hospitals with average patient age under 40

How to get started | **Data Driven Approach**



Let’s teach machines

- Demo

Project A

Uses ML and patient Records to predict negative patient outcomes

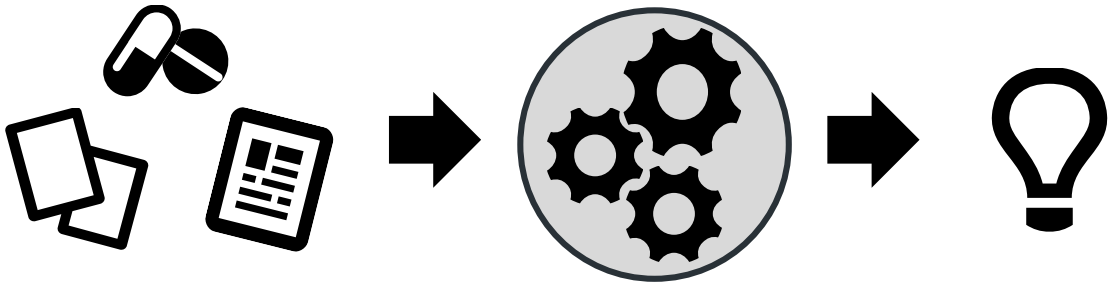
Project B

- Uses user-feedback to innovatively improve the ML Model

Project C

- Grass roots ML project that yielded insight into factors that affect the length of a patients stay.

Machine Learning in Healthcare | **Project A**



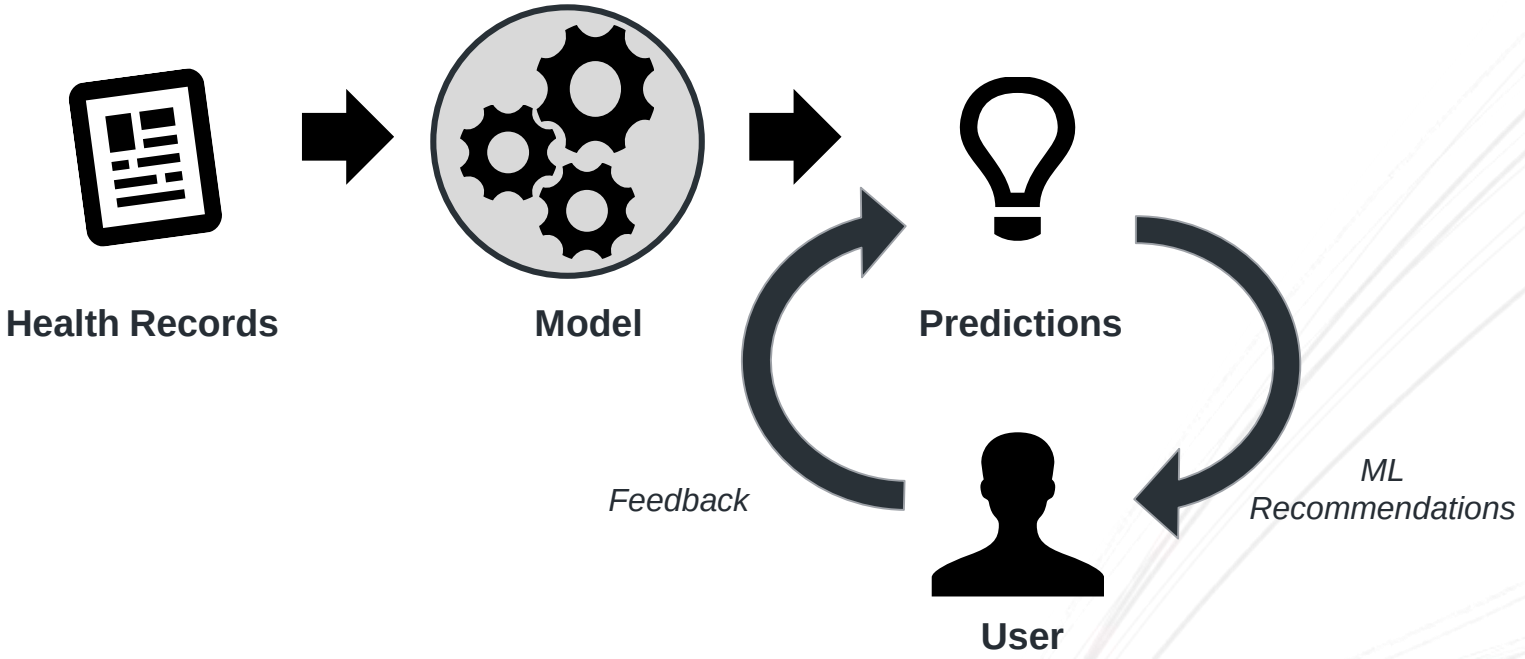
Data (Pt EMR):
Meds
Labs
Visit History
Notes

Model

Predictions
+ SEPSIS
+ Adverse Drug Reaction
+ % Readmission

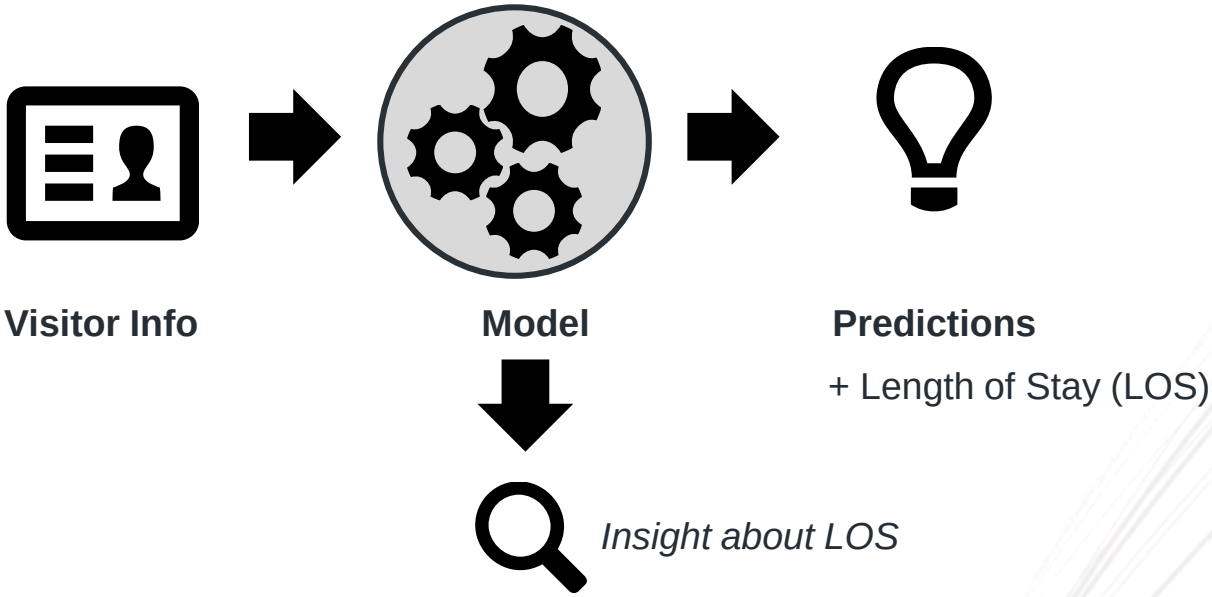
- Hospitals already have digital records, because of the ACA
- Can use public algorithm and software to make predictions
- Predictions save lives and/or money

Machine Learning in Healthcare | **Project B**



- System uses proprietary model to make recommendations to clinicians
- Hospital decided to improve system by developing a way to incorporate feedback
- Hospital improves the model overtime without having to buy a new system

Machine Learning in Healthcare | **Project C**



- Interested individuals built a model using visit data
- Predictions were just ***OKAY***
- ***BUT*** inspecting the model allowed the team to learn new factors that were unknown about the different factors affected the clincial outcomes

Questions

References

Resource	Location
Presentation link	
EDX Course list (see data)	https://www.edx.org/course
Coursera Course list	https://www.coursera.org/courses/?domains=data-science
Spark Download Page	http://spark.apache.org/downloads.html
Anaconda Python Install Page	http://docs.continuum.io/anaconda/install
R install page	https://cran.r-project.org/
R Studio install page	https://www.rstudio.com/products/rstudio/download/
Ebay Tech Blog on Spark	http://www.ebaytechblog.com/2014/05/28/using-spark-to-ignite-data-analytics/