



# MLADS

MACHINE LEARNING, AI,  
AND DATA SCIENCE CONFERENCE

JUNE, 2020



# Generalization, Utility, and Experimentation: ML Concepts for Making Better Business Decisions

Robert Horton  
John-Mark Agosta  
Mario Inchiosa

# Session goals

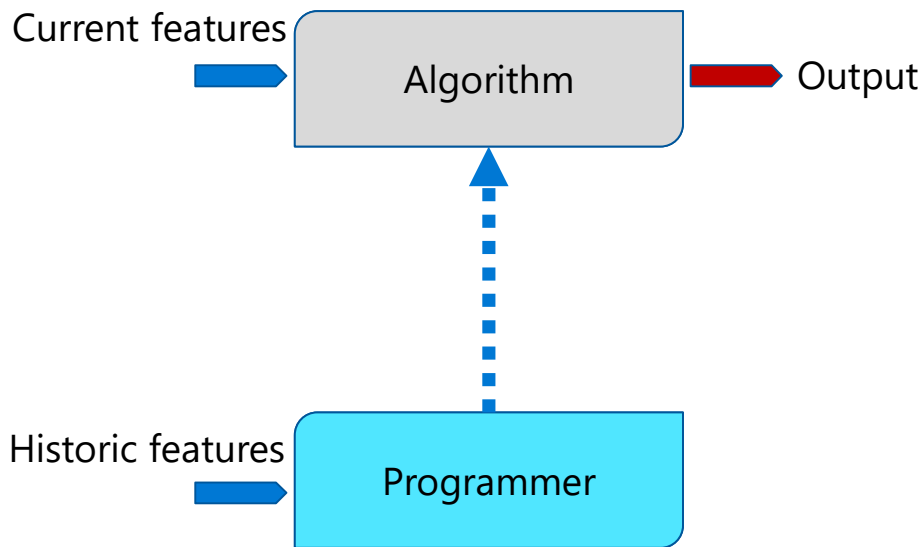
1. **Learn how machine learning (ML) differs** from traditional software engineering
2. See how ML fits in the context of **making better business decisions**
3. Understand why causal relationships matter in data analysis, and **why we still need to do experiments**



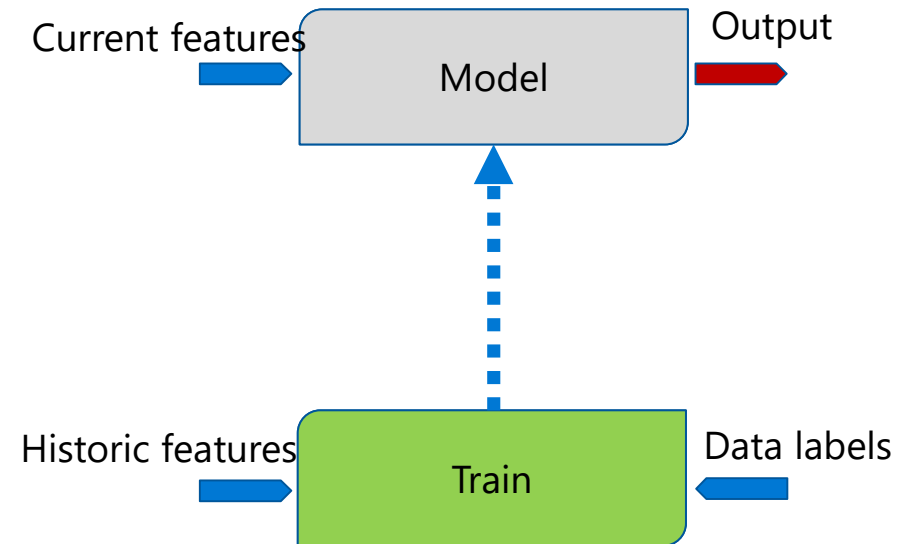
# Learning vs. Coding: Building software systems that manage probability

John-Mark Agosta

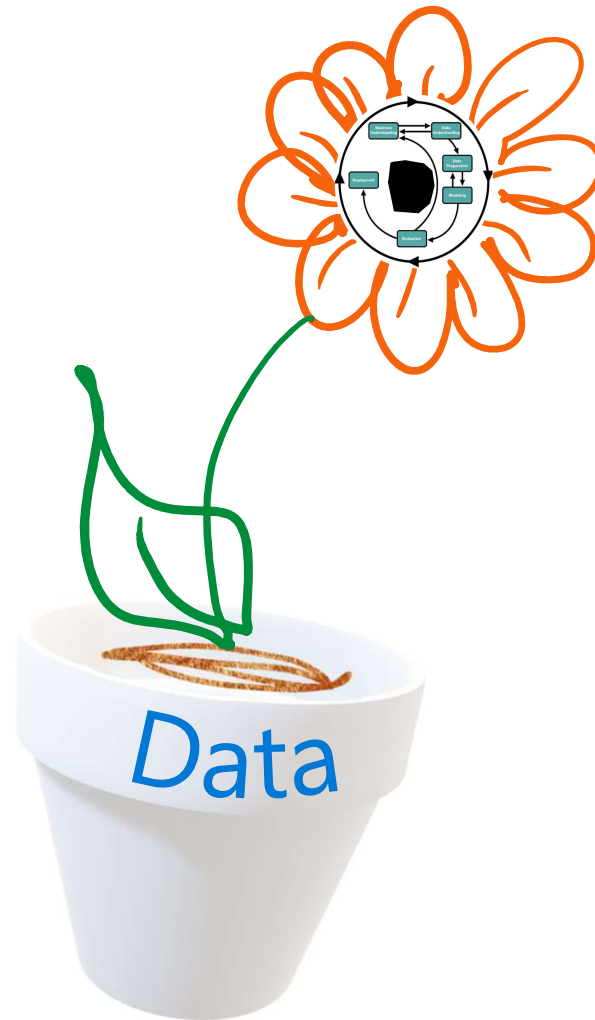
# Manual programming v/s Machine Learning



In **conventional programming** the programmer comes up with the algorithm



In Supervised **Machine Learning**, a statistical model is generated from labels applied to historic data



# Expert Intuition v/s Statistical Models

One of the most durable social science findings is that simple statistical rules outperform human expert judgment consistently across many fields, when predicting long term outcomes. (Paul Meehl's little book: *Clinical versus Statistical Prediction* (1954).)

- Typically only a concise number of factors are necessary – how are these discovered? (This goes to questions about causality)
- *How do you determine if your problem is one where such ML/statistical methods apply?*

## DEMO: Crowd-sourced rules v/s Automated ML

Imagine you need a set of decision rules that distinguish messages from different newsgroups (circa 1993) How do your hand-crafted rules compare to a machine-learning generated classifier?

*Try this demo and see.*

1. Go to <http://aka.ms/news-rules>
2. Rinse and repeat!



# Automated Learning of Behavior from Data

- ML combines many weak correlations
  - when hand-coded Rules fail
- By optimizing for accuracy over labeled examples
  - It Approximates and Generalizes
- Minimizing the residual probability of error
  - Take advantage of Uncertainty to improve top-level performance

*"Unpredictable consequences are the most expected thing on earth."*

--- B. Latour

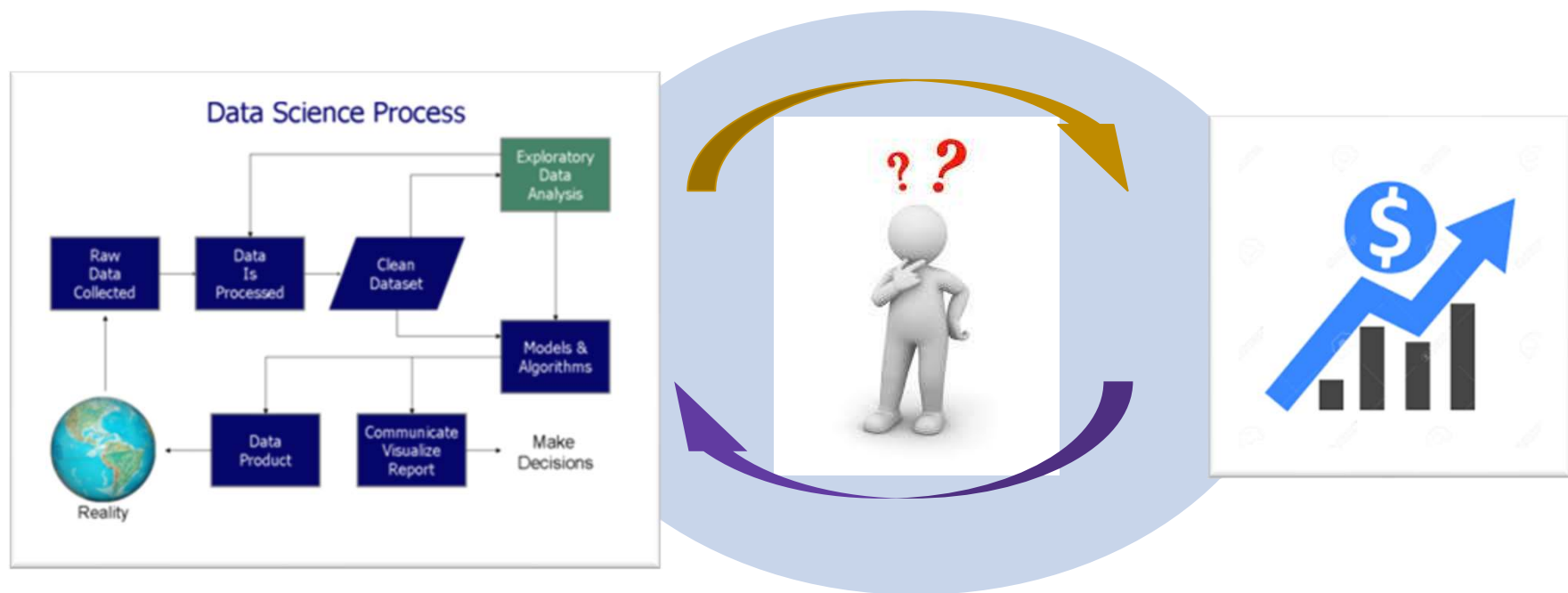
# Confusion Matrix to measure Accuracy

A confusion matrix compares the outcome of a prediction to the actual, observed values by tabulating the counts where they agree and disagree.

observed predicted	True	False
True	TP	FP
False	FN	TN

TP: "True Positive", TN: "True Negative"  
FP: "False Positive", FN: "False Negative"

**STRATEGY: The Data Science Process can begin only after you've grasped the Decision Problem.**



*As a Data Scientist, where do you start?*

# Automating Decisions with ML:

## Success depends on the *Decision Quality*.

1. Solve the right problem.
2. Consider a true range of alternatives.
3. Search out curated and relevant information,
4. Value consequences of decision outcomes,
5. Apply solid reasoning.
6. Implement what has been decided.

