



2018 Top 10

Business Intelligence Trends

Don't Fear Artificial Intelligence

1.
Don't Fear Artificial Intelligence

2.
Liberal Arts Impact

3.
Promise of NLP

4.
Multi-Cloud Debate

5.
Rise of the Chief Data Officer

6.
Crowdsourced Governance

7.
Data Insurance

8.
Data Engineer Role

9.
Location Internet of Things

10.
Academics Investment



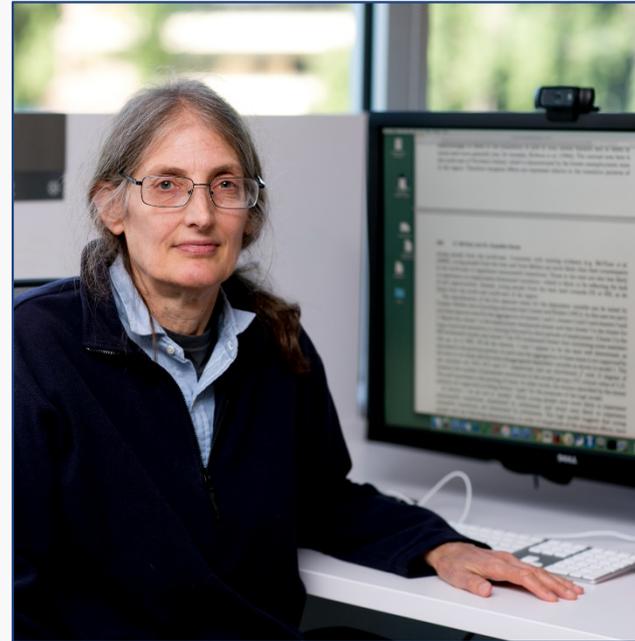
Don't Fear Artificial Intelligence

Moderator & Speakers



Andy Cotgreave
Technical Evangelism
Director

Moderator



Chris Fraley
Senior Research
Scientist



Eric Brochu
Senior Software
Engineer

Who we are



Andy Cotgreave

Andy is a visual analytics expert who has been with Tableau since 2011 in various roles ranging from product consultant to social content manager. He is now Tableau's technical evangelism director. Prior to Tableau, he was a data analyst at the University of Oxford.

As a technical evangelist, Andy helps people see and understand their data using Tableau's innovative products. He shares his passion for visual analysis and technology with his writing, (e.g. Computerworld, on tableau.com, and his own blog), speaking at industry conferences like SXSW and Tableau's own events.

Who we are



Eric Brochu

- PhD in CS from University of British Columbia
 - Thesis in Machine Learning, worked on models for nonlinear optimization problems using human input
- Joined Tableau in March 2016
 - Working on Recommendations team, helping design and implement recommender systems for tables & joins and data sources.

Who we are



Chris Fraley

- PhD in CS from Stanford University
 - 25+ years experience in statistical computing
- Joined Tableau in November 2015
 - Working on Analytics research and development unit
 - Contributed to Tableau's clustering algorithm

Agenda

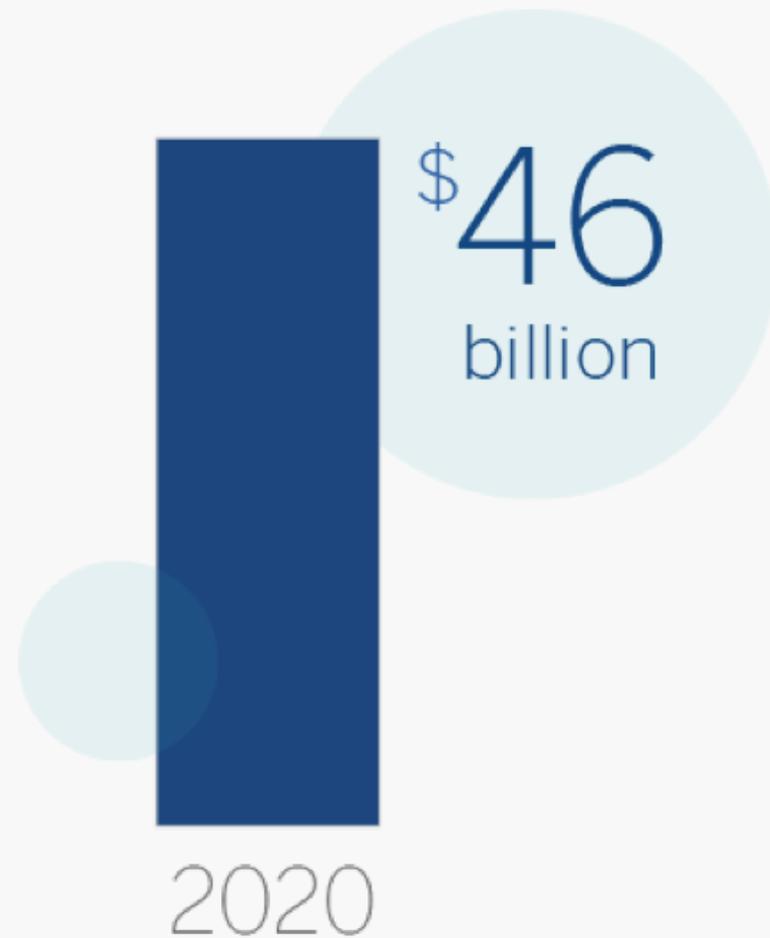
1. Don't Fear Artificial Intelligence
2. Machine Learning Definitions
3. Advantages & Limitations with Machine Learning
4. Machine Learning at Tableau
5. Open Discussion with Panel



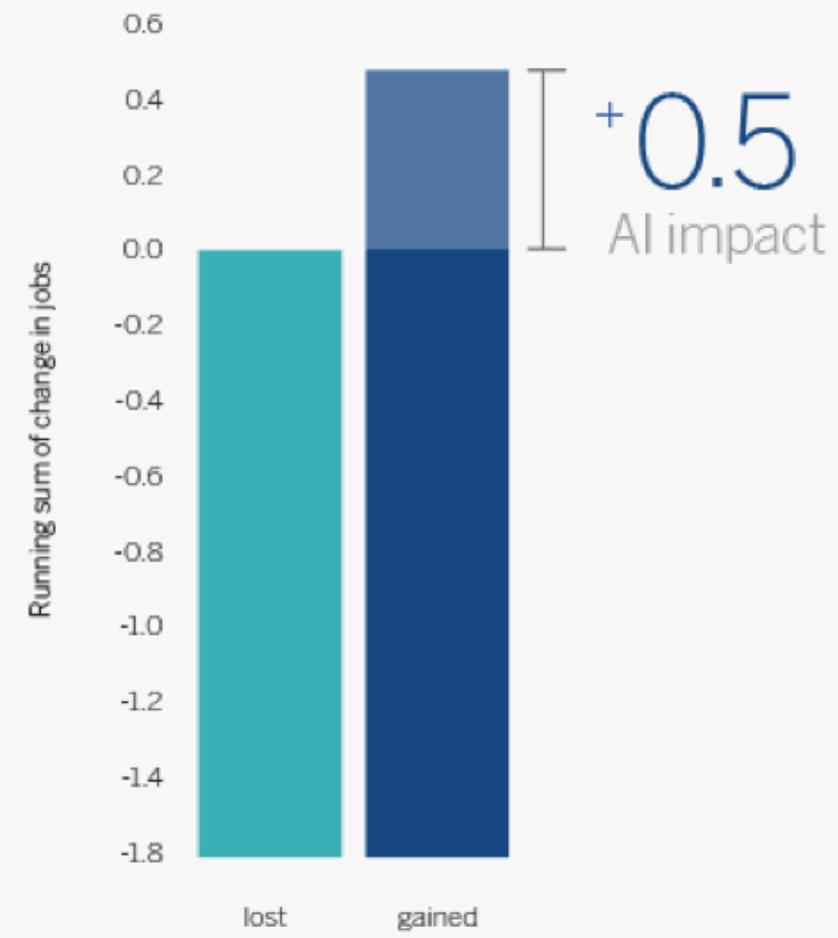
Machine Learning Introduction

AI & Machine Learning To Boom

IDC forecasts revenues from AI and machine learning systems to total \$46 billion by 2020



In 2020, AI will become a positive net job motivator, creating 2.3 million jobs, Gartner reports



“The development of full artificial intelligence could spell the end of the human race.”

- *Stephen Hawking*
(*BBC Interview, 2014*)



“The development of full artificial intelligence **could** spell the end of the human race.”

- *Stephen Hawking*
(*BBC Interview, 2014*)



“I don't work on not turning AI evil today for the same reason I don't worry about the problem of overpopulation on the planet Mars.”

*Andrew Ng,
VP & Chief Scientist, Baidu*



AlphaGo Zero

Starting from scratch



AlphaGo Zero

Starting from scratch

'Read this and embrace the future' WALTER ISAACSON

Deep Thinking

Where Machine Intelligence Ends



And Human Creativity Begins

Garry Kasparov

'Optimistic, wise and compelling' CHARLES DUHIGG

Home > Technology

Robots will take a third of British jobs by 2030, report says



The Starship delivery robot has already started transporting food orders

News > Science

The robots are coming – but will they really take all our jobs?

Artificial intelligence is going to transform the world, and our lives. But are we heading for a brave new world, or a science fiction horror-show?

David Barnett | Thursday 30 November 2017 06:15 GMT | 58 comments

f t ✉ 168 shares

Like Click to follow The Independent Online



Menial, repetitive drudgery could be absorbed by robots, which would leave us humans more free time – but to do what? Getty

Last week, Chancellor Philip Hammond announced in the Autumn Budget a £500m package of investment into tech

PROMOTED STORIES

News > Science

The robots are coming and they will really take all our jobs

Artificial intelligence is going to transform our new world, or a science fiction horror-show

David Barnett | Thursday 30 November 2017 06:15 GMT | 58

168 shares



Menial, repetitive drudgery could be absorbed by robots, which

Last week, Chancellor Philip Hammond announced in the Autumn Budget a £500m package of investment into tech

Up to 30% of existing UK jobs could be impacted by automation by early 2030s, but this should be offset by job gains elsewhere in economy

Mar 24, 2017

- Up to around 30% of existing UK jobs are susceptible to automation from robotics and Artificial Intelligence (AI) by the early 2030s, but in many cases the nature of jobs will change rather than disappear

PROMOTED STORIES

British jobs by



Supporting food orders



Machine Learning Definitions

What is Machine Learning?



What is Machine Learning?

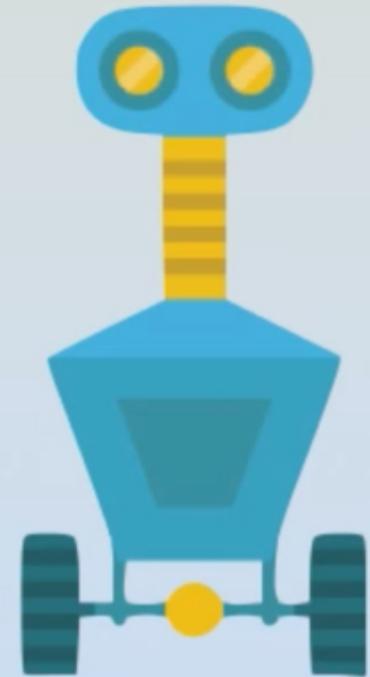
Learn from experience



Learn from ~~experience~~ ^{data}



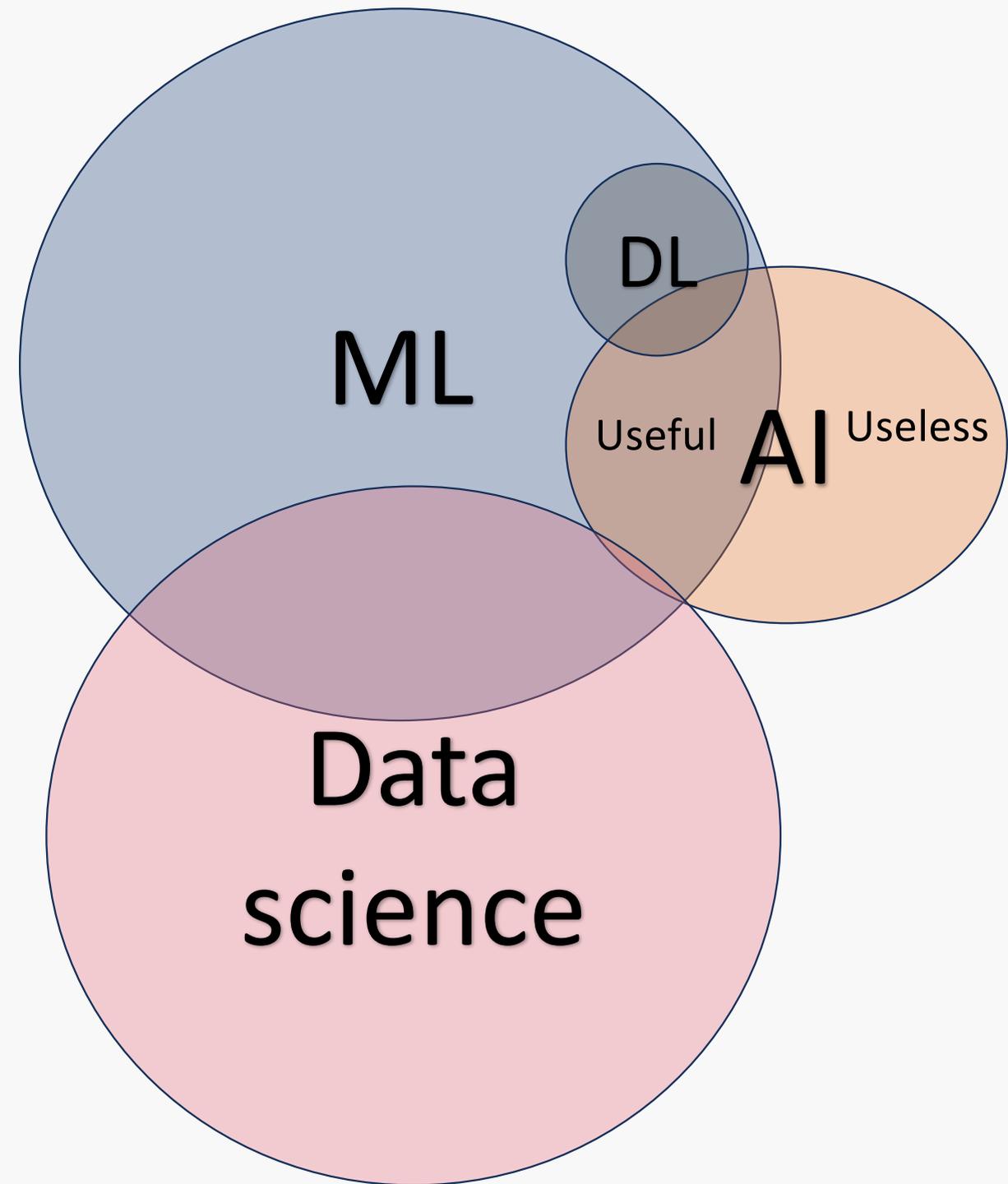
Follow instructions



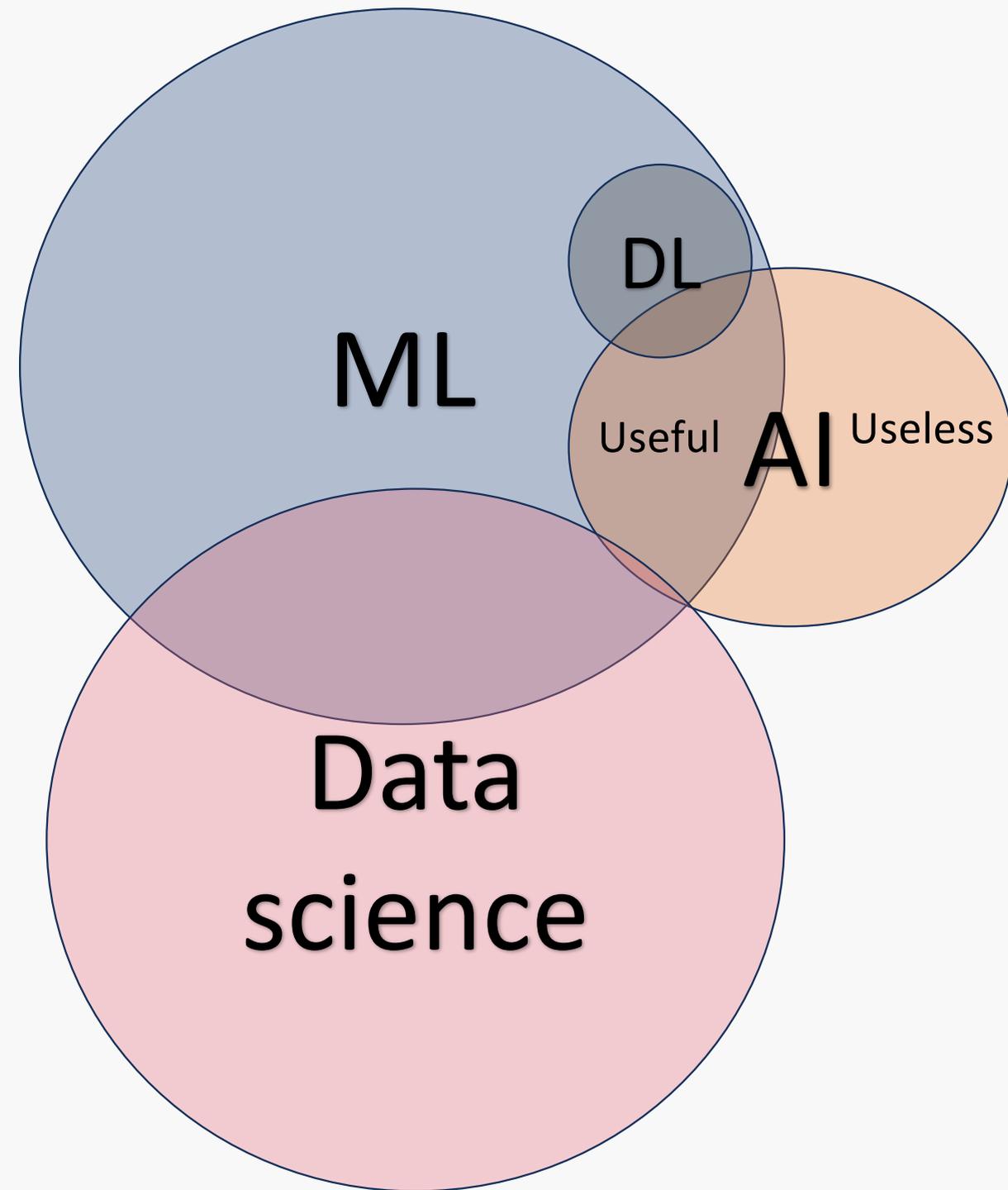
How does it differ from AI? Data science? Deep learning?



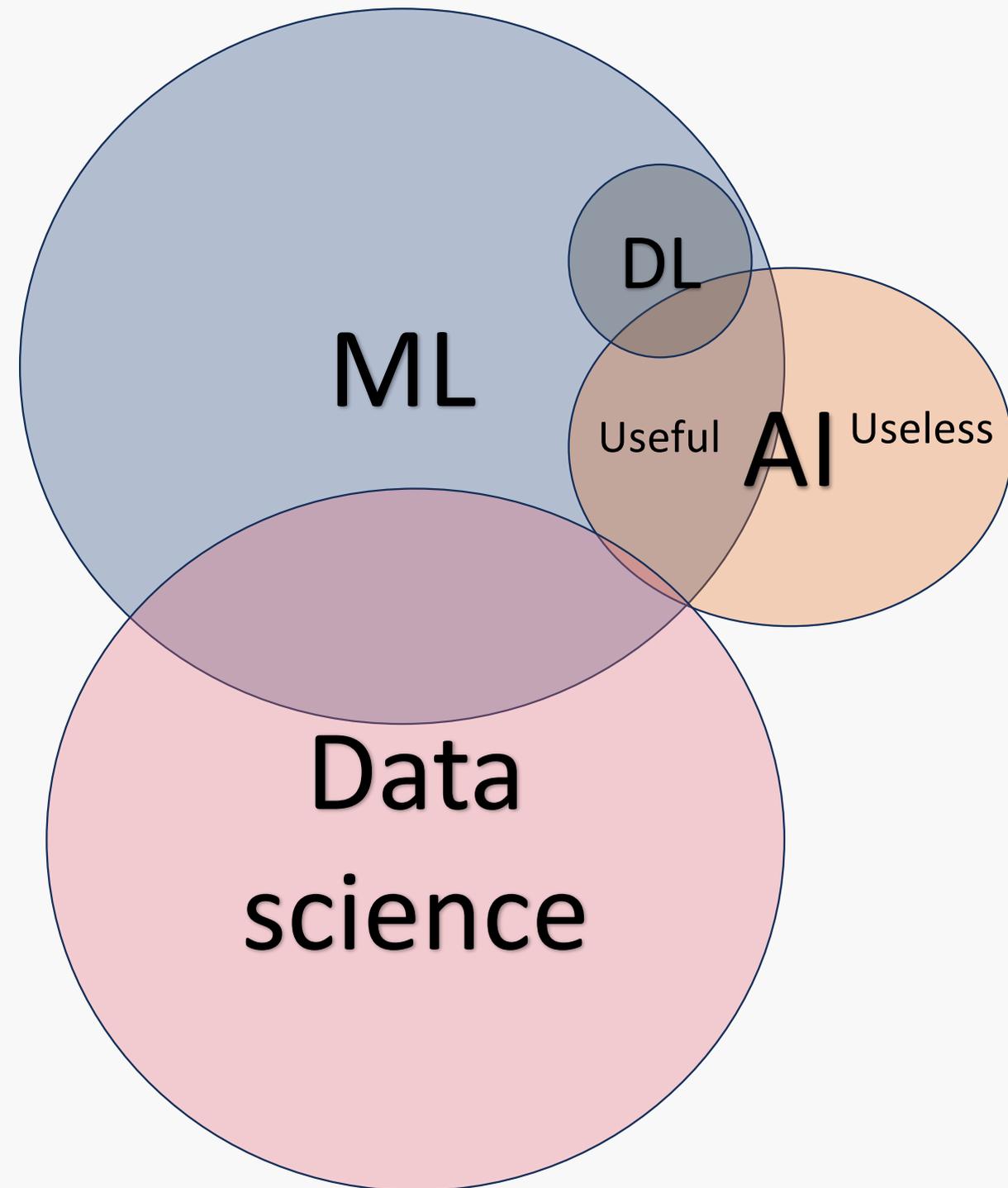
- Machine learning
 - Methods which are trained on past data to produce future output



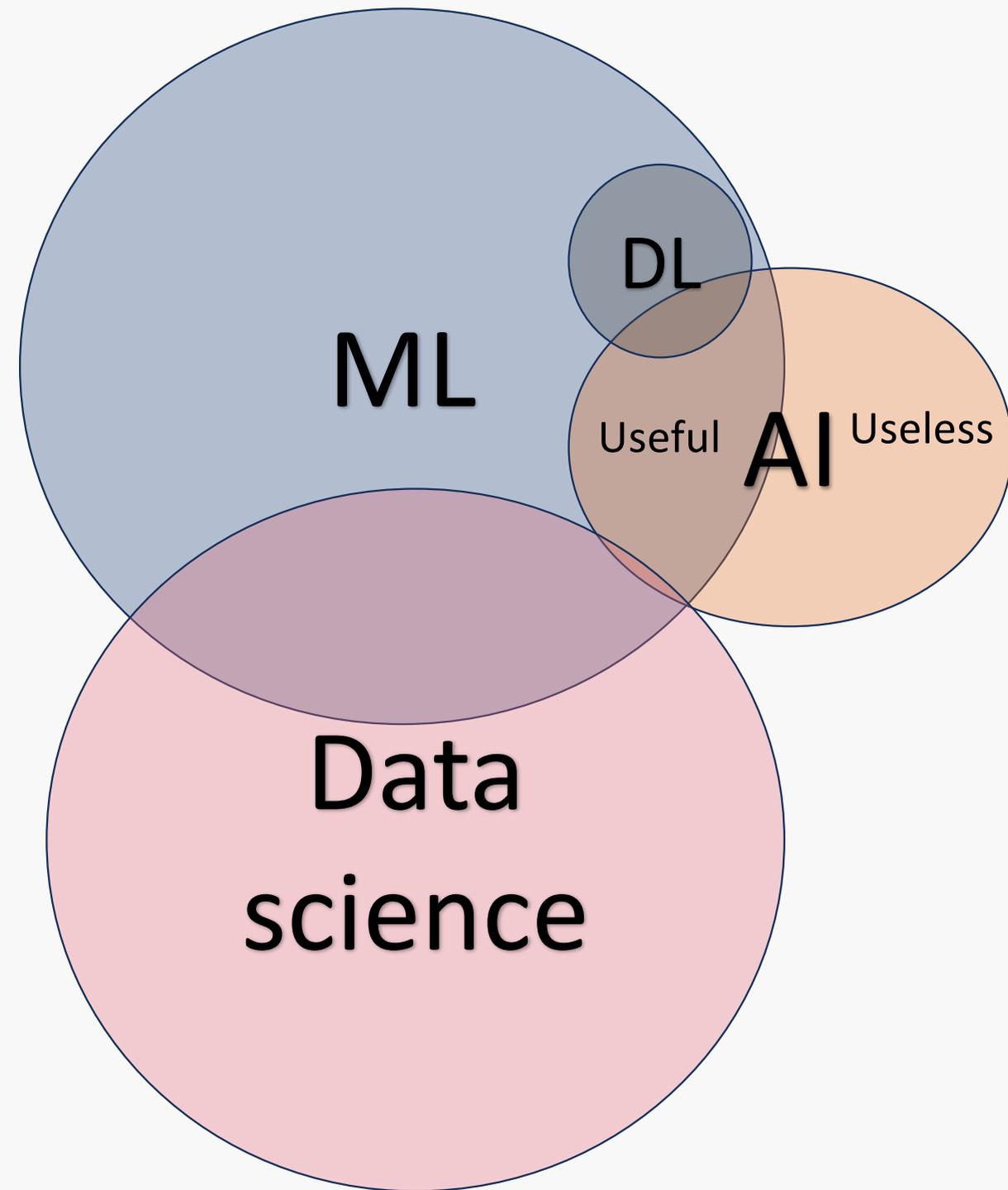
- Machine learning
 - Methods which are trained on past data to produce future output
- AI
 - Mimicking *human* understanding and behavior
 - No requirement to use ML
 - ML happens to work the best



- Machine learning
 - Methods which are trained on past data to produce future output
- Deep learning
 - Specific collection of Neural Net based ML methods
 - Pretty amazing results on certain problems



- Machine learning
 - Methods which are trained on past data to produce future output
- Data science
 - Full pipeline to solve problems with data
 - Often uses ML, focuses less on automation
- Statistics
 - Same / similar techniques
 - Usually different goals

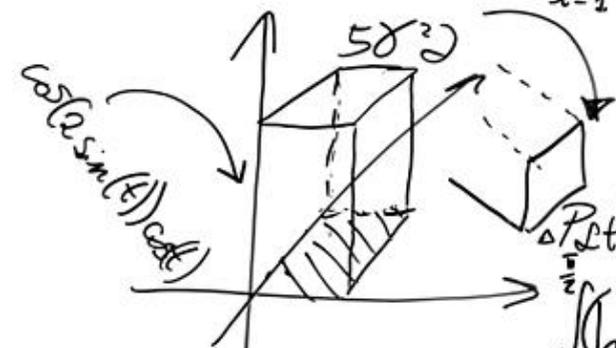




Advantages / Limitations of ML

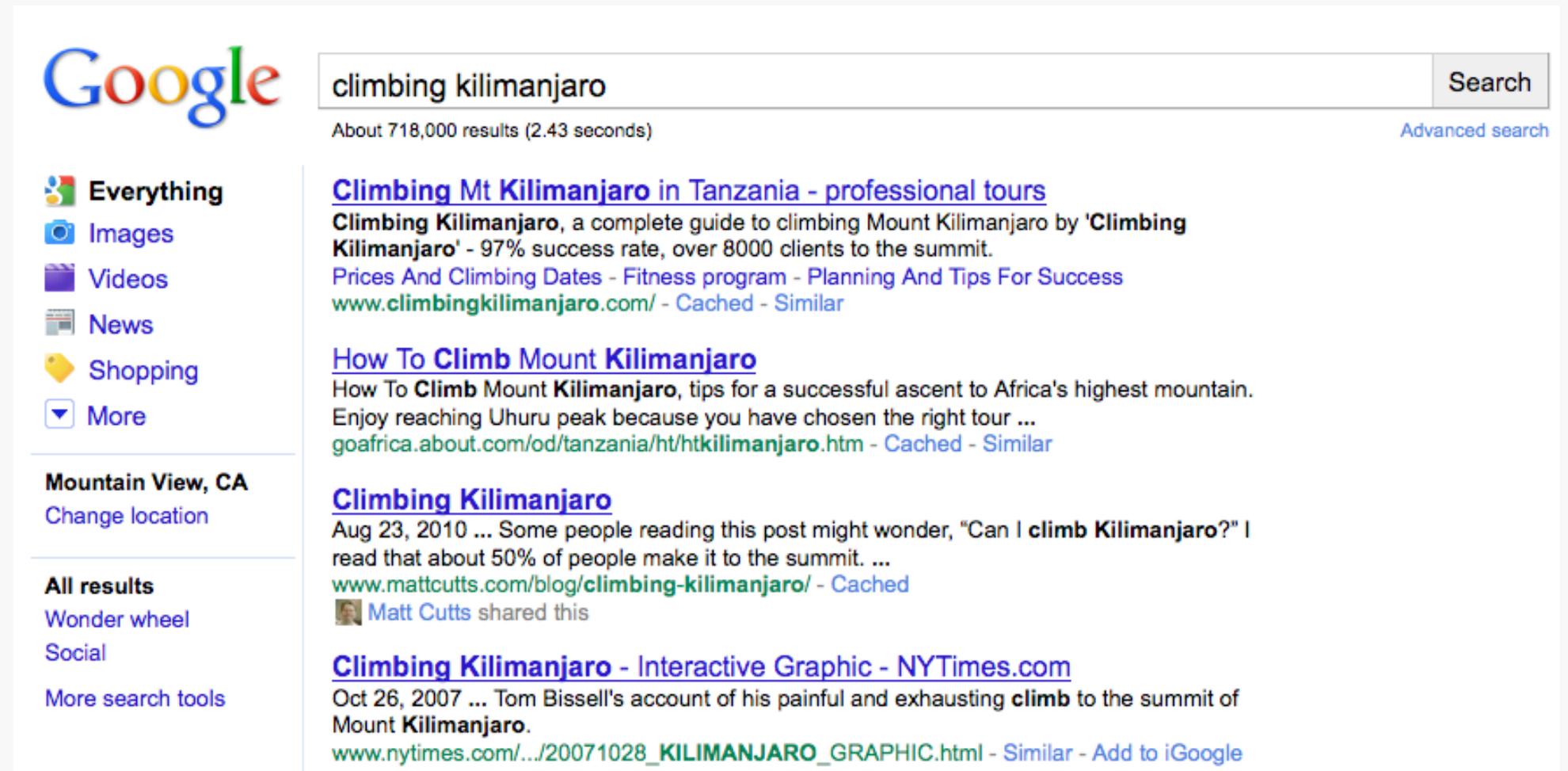
Advantages of Machine Learning

#1: Don't have to develop math/stat algorithms from scratch

$\mathcal{L} = \oint E \cdot dt$
 $f(\omega) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \omega} dx \frac{dt}{d\omega}$
 $\rho \left(\frac{\partial v}{\partial t} + v \cdot \nabla v \right) = -\nabla p + \nabla \cdot T + f$
 $\nabla \cdot E = 0$
 $\nabla \times E = -\frac{1}{c} \frac{\partial H}{\partial t}$
 $\nabla \cdot H = 0$
 $\nabla \times H = \frac{1}{c} \frac{\partial E}{\partial t}$
 $-i\hbar \frac{\partial}{\partial t} \Psi = H \Psi$
 $H = -\sum p(x) \log p(x)$
 $\frac{1}{2} G^2 S^2 \frac{\partial^2 V}{\partial S^2} + r S \frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} - r \cdot V = 0$
 $+ \sum_{i=1}^n \frac{q_i}{2} H_i^M + c_s \frac{D}{Q} + c_o D + \frac{Q(p-D)}{2p} M^M + F_o N + F_o N + \sum_{i=1}^n D_i \omega_i d_i \frac{(1+\omega_i)}{F_i}$
 $TC(Q, q_i, m_i) = \sum_{i=1}^n \left[\frac{D_i}{m_i q_i} S_i + c_i^v D_i + \frac{q_i H_i^v}{2} \left(m_i \left(1 - \frac{D_i}{P_i} \right) - 1 + 2 \frac{D_i}{P_i} \right) \right] +$

 $\left[\frac{d \Delta p(s, \phi)}{d\phi} \right] = \begin{bmatrix} \beta & -\beta \\ -\beta & 0 \end{bmatrix} \begin{bmatrix} \Delta p(s, \phi) \\ \Delta M(s, \phi) \end{bmatrix}$
 $\int_0^{\frac{\pi}{2}} (\log \sin x)^2 dx = \int_0^{\frac{\pi}{2}} (\log \cos x)^2 dx = \frac{\pi}{2} \left\{ \frac{\pi^2}{12} + (\log 2)^2 \right\}$

Advantages of Machine Learning

#2: Speed up decision making through automation

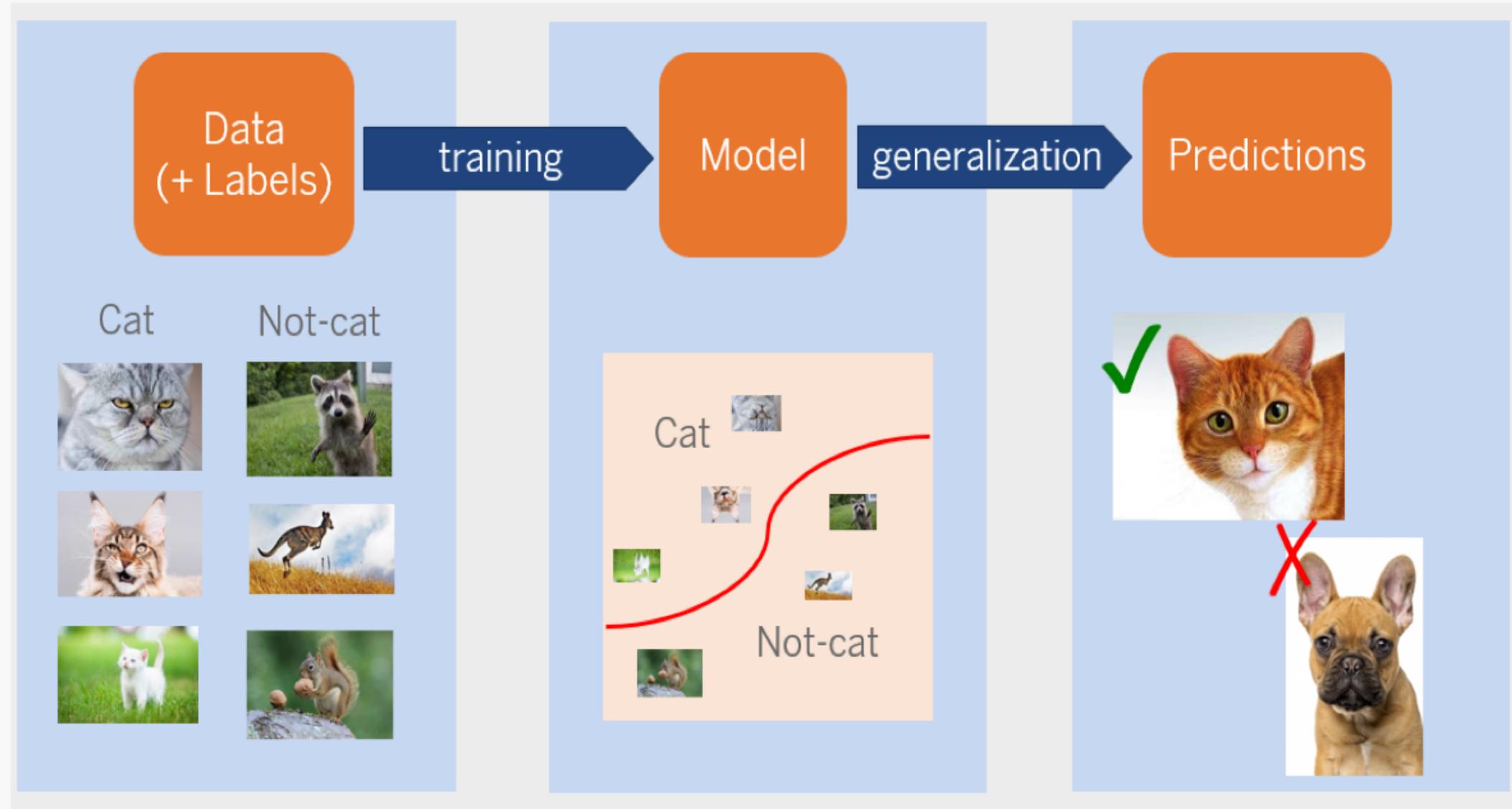


The image shows a Google search interface for the query "climbing kilimanjaro". The search bar is at the top right, with the text "climbing kilimanjaro" and a "Search" button. Below the search bar, it says "About 718,000 results (2.43 seconds)" and "Advanced search". On the left side, there are navigation options: "Everything", "Images", "Videos", "News", "Shopping", and "More". Below these, it shows the location "Mountain View, CA" and "Change location". Under "All results", there are links for "Wonder wheel", "Social", and "More search tools". The search results are listed on the right side, with the following entries:

- Climbing Mt Kilimanjaro in Tanzania - professional tours**
Climbing Kilimanjaro, a complete guide to climbing Mount Kilimanjaro by 'Climbing Kilimanjaro' - 97% success rate, over 8000 clients to the summit.
Prices And Climbing Dates - Fitness program - Planning And Tips For Success
www.climbingkilimanjaro.com/ - Cached - Similar
- How To Climb Mount Kilimanjaro**
How To Climb Mount Kilimanjaro, tips for a successful ascent to Africa's highest mountain. Enjoy reaching Uhuru peak because you have chosen the right tour ...
goafrica.about.com/od/tanzania/ht/htkilimanjaro.htm - Cached - Similar
- Climbing Kilimanjaro**
Aug 23, 2010 ... Some people reading this post might wonder, "Can I climb Kilimanjaro?" I read that about 50% of people make it to the summit. ...
www.mattcutts.com/blog/climbing-kilimanjaro/ - Cached
Matt Cutts shared this
- Climbing Kilimanjaro - Interactive Graphic - NYTimes.com**
Oct 26, 2007 ... Tom Bissell's account of his painful and exhausting climb to the summit of Mount Kilimanjaro.
www.nytimes.com/.../20071028_KILIMANJARO_GRAPHIC.html - Similar - Add to iGoogle

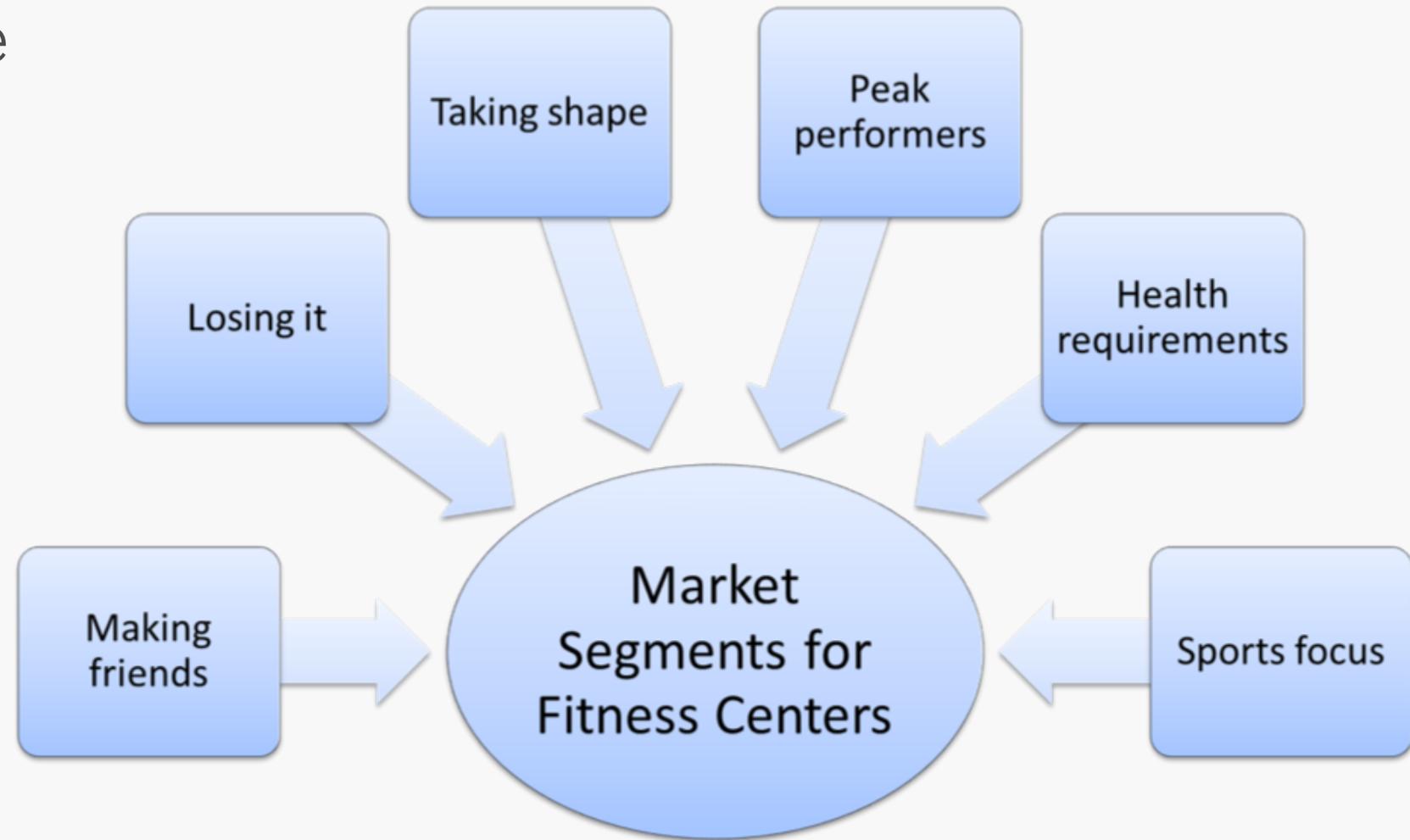
Advantages of Machine Learning

#3: Learn automatically through examples (e.g. classification)



Advantages of Machine Learning

#4: Discover patterns, like clustering/segmentation or outlier/anomaly detection



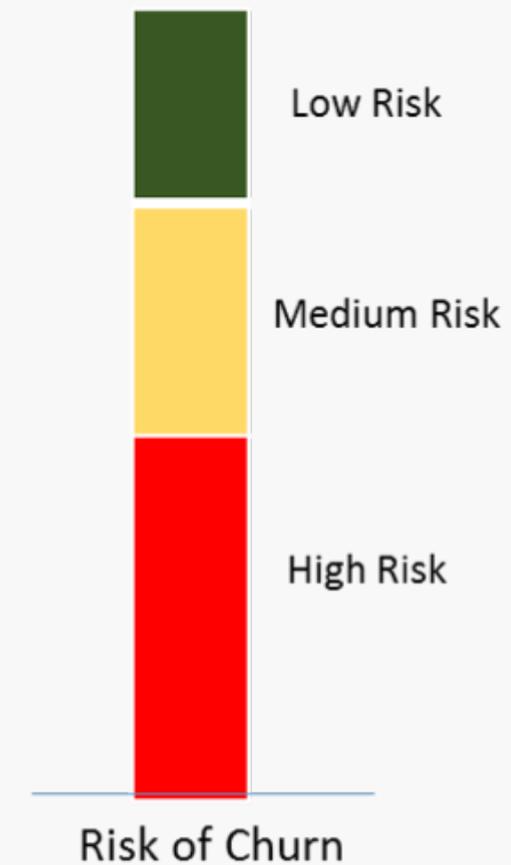
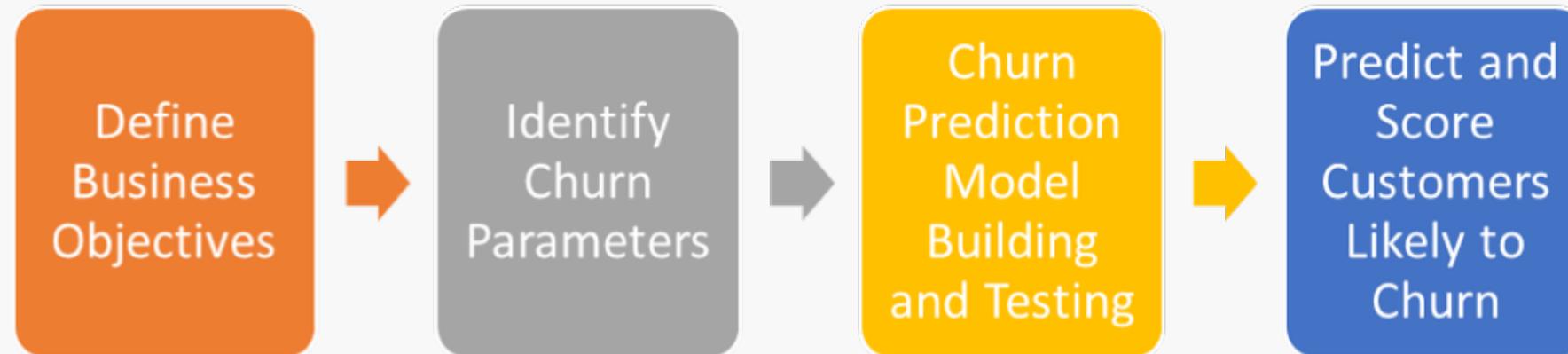
Advantages of Machine Learning

#4: Discover patterns, like clustering/segmentation or outlier/anomaly detection



Advantages of Machine Learning

#5: Predict outcomes



Advantages of Machine Learning

#6: Identify which inputs most affect outcomes

Customer loyalty in telecom operators

COMARCH CRM & MARKETING

Customer churn – why do customers change operators?

The top 3 reasons why subscribers change providers:



- ! They want a new handset
- ! They believe they pay too much for calls
- ! Providers do not offer additional loyalty benefits

Source: Research conducted by Analysys Mason and Buongiorno

Kraków/ 09.08.2012

Page 8

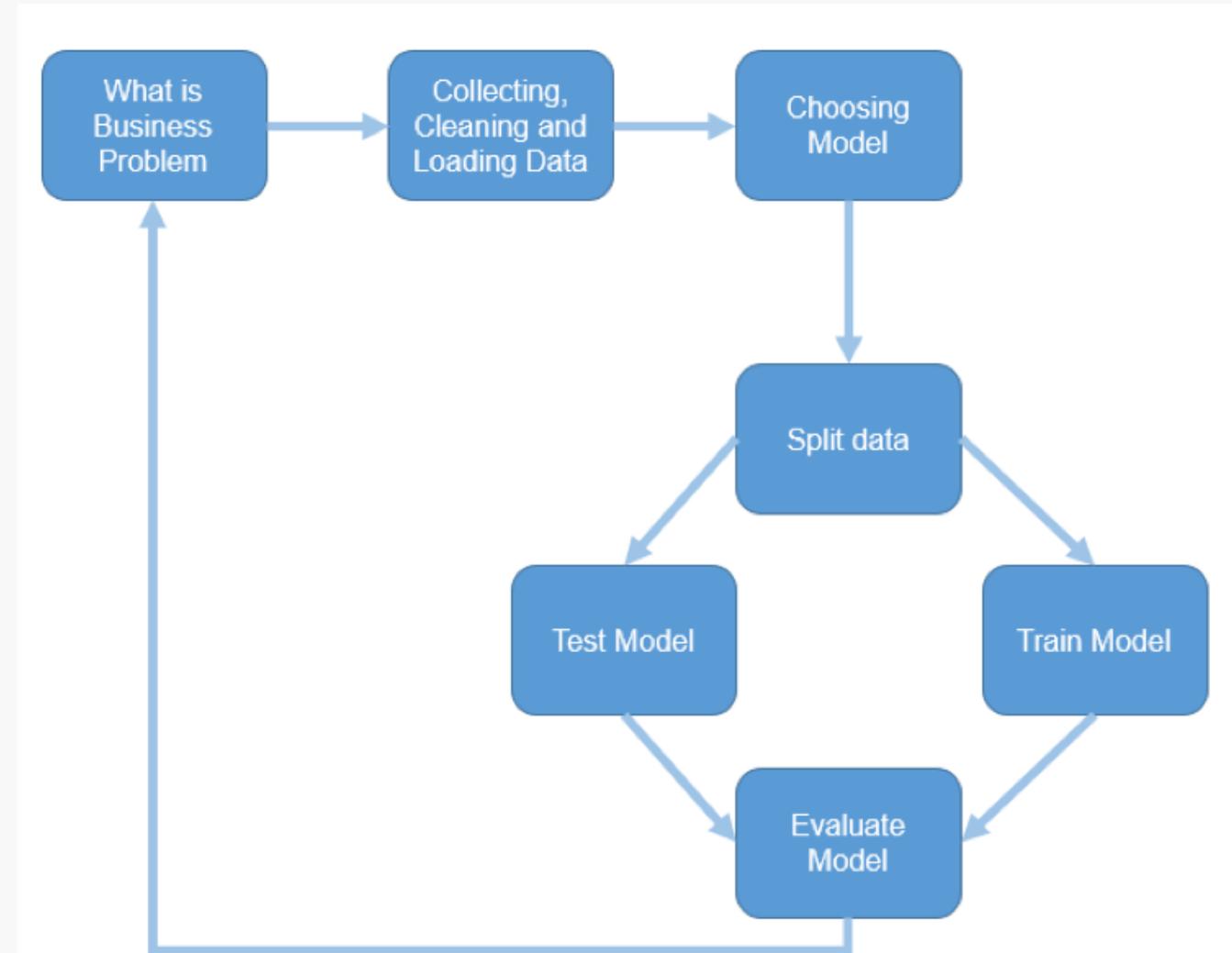
Limitations of Machine Learning

#1: Doesn't eliminate the importance of quantity and quality of data collected



Limitations of Machine Learning

#2: Need for validation and uncertainty quantification



Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

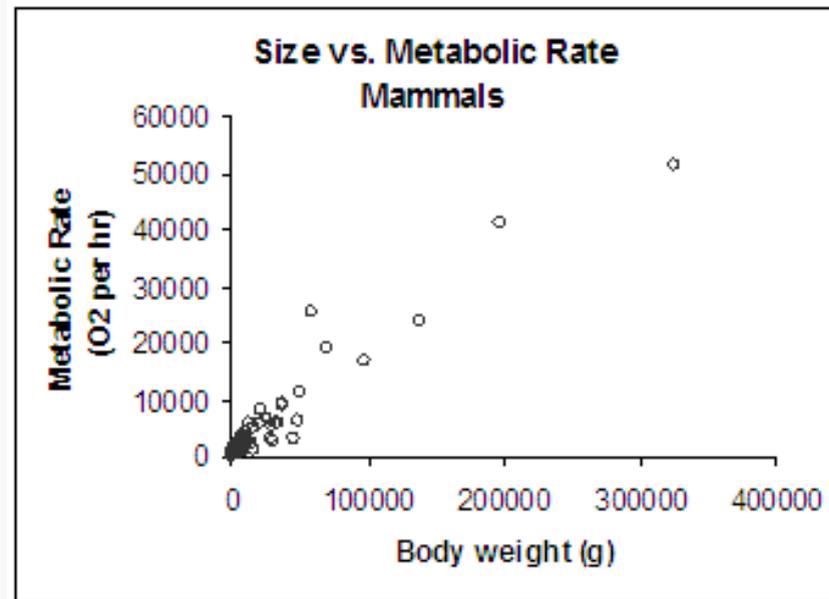
Model selection

Variable selection

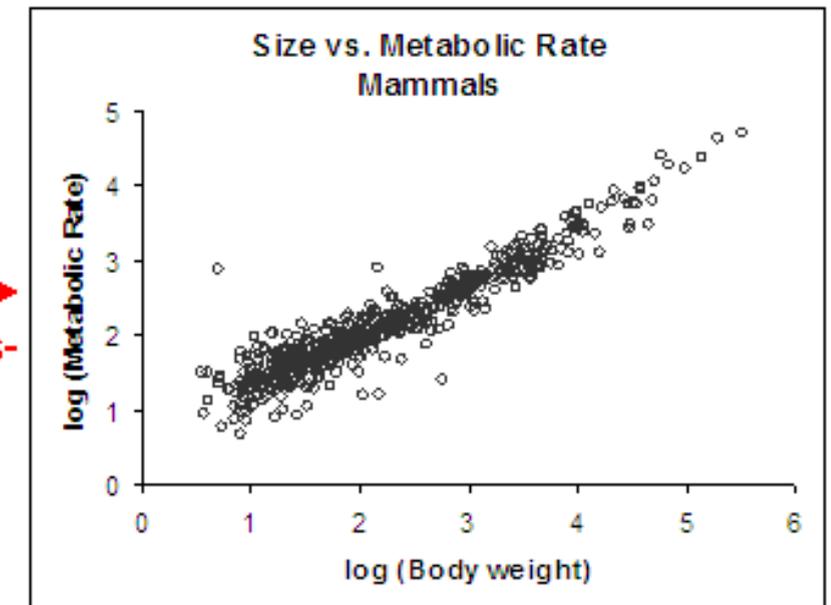
Parameter settings

Handling missing data

Handling outliers/anomalies



Log
→
Transform



Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

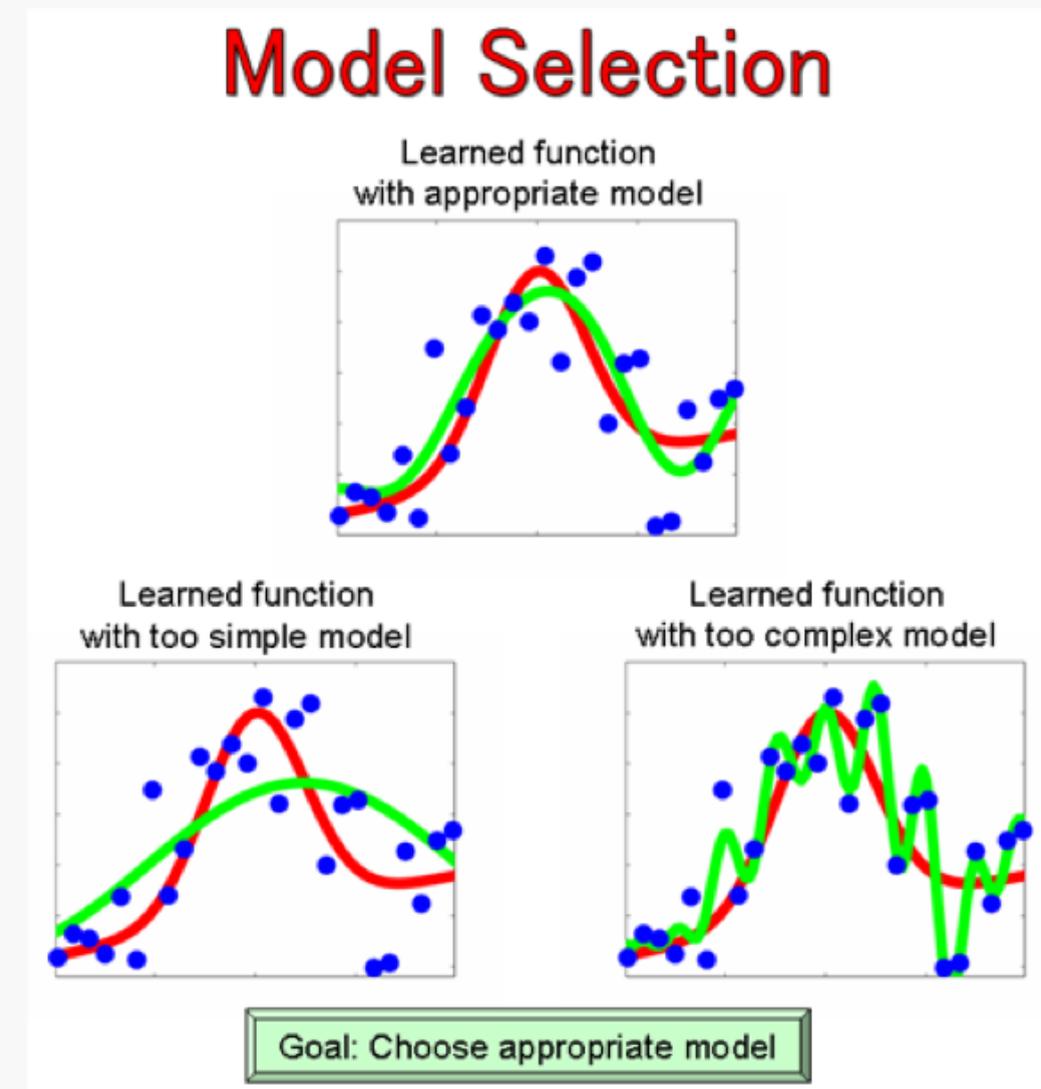
Model selection

Variable selection

Parameter settings

Handling missing data

Handling outliers/anomalies



Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

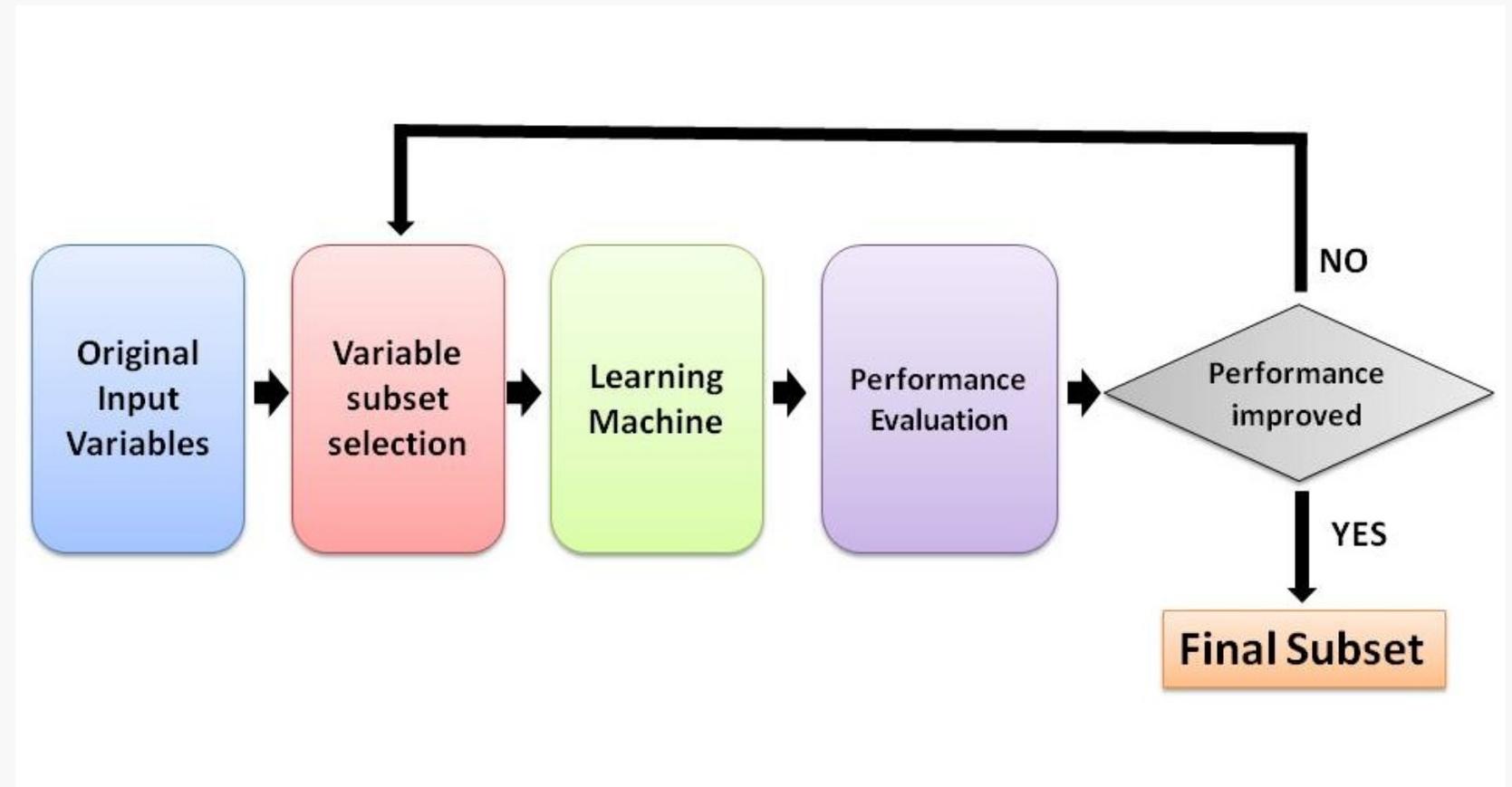
Model selection

Variable selection

Parameter settings

Handling missing data

Handling outliers/anomalies



Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

Model selection

Variable selection

Parameter settings

Handling missing data

Handling outliers/anomalies

Hyper-parameter	Variants
Non-linearity	linear, tanh, sigmoid, ReLU, VReLU, RReLU, PReLU, ELU, maxout, APL, combination
Batch Normalization (BN)	before non-linearity. after non-linearity
BN + non-linearity	linear, tanh, sigmoid, ReLU, VReLU, RReLU, PReLU, ELU, maxout
Pooling	max, average, stochastic, max+average, strided convolution
Pooling window size	3x3, 2x2, 3x3 with zero-padding
Learning rate decay policy	step, square, square root, linear
Colorspace & Pre-processing	RGB, HSV, YCrCb, grayscale, learned, CLAHE, histogram equalized
Classifier design	pooling-FC-FC-clf, SPP-FC-FC-clf, pooling-conv-conv-clf-avepool, pooling-conv-conv-avepool-clf
Network width	$1/4$, $1/2\sqrt{2}$, $1/2$, $1/\sqrt{2}$, 1 , $\sqrt{2}$, 2 , $2\sqrt{2}$, 4 , $4\sqrt{2}$
Input image size	64, 96, 128, 180, 224
Dataset size	200K, 400K, 600K, 800K, 1200K(full)
Batch size	1, 32, 64, 128, 256, 512, 1024
Percentage of noisy data	0, 5%, 10%, 15%, 32%
Using bias	yes/no

Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

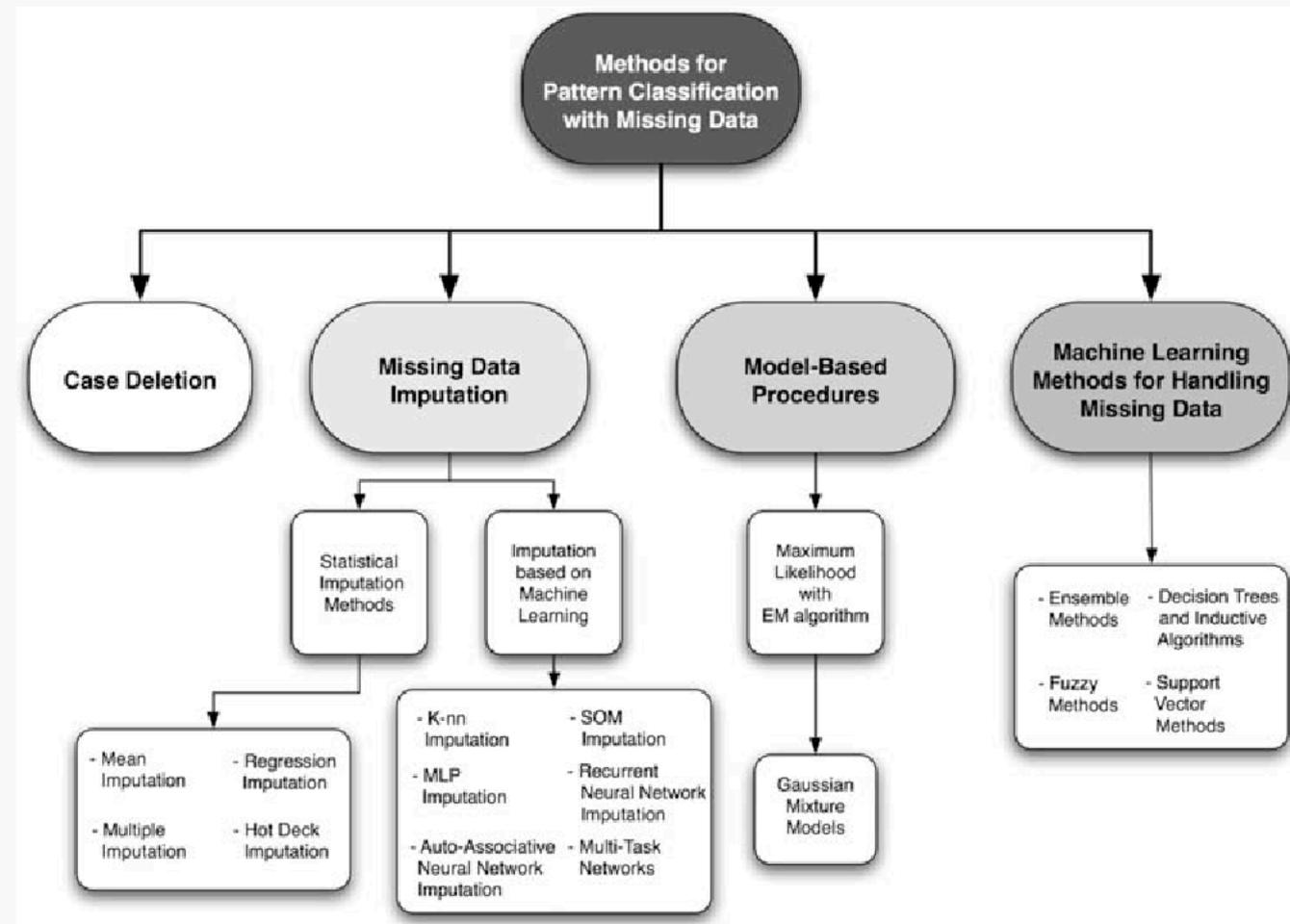
Model selection

Variable selection

Parameter settings

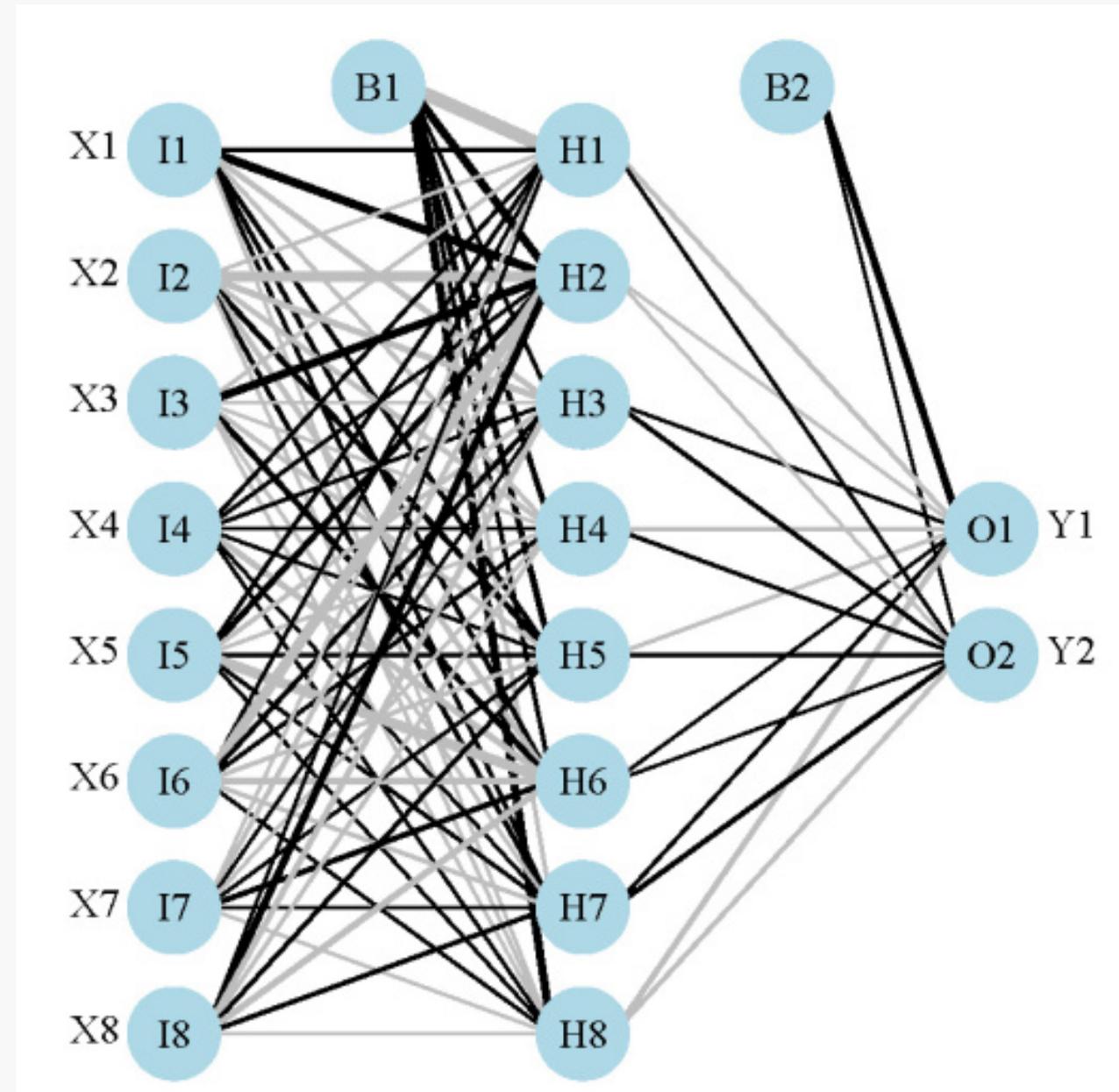
Handling missing data

Handling outliers/anomalies



Limitations of Machine Learning

#4: Machine Learning models can be hard to interpret





Machine Learning at Tableau

Smart @ Tableau

Human-Centered

- Augmentative.
- Empowering.
- Focus on things machines are good at and humans are not.
- Don't interrupt me!

Contextual

- What is the task?
- Knowledge elicitation.
- What to infer and how?

Adaptive

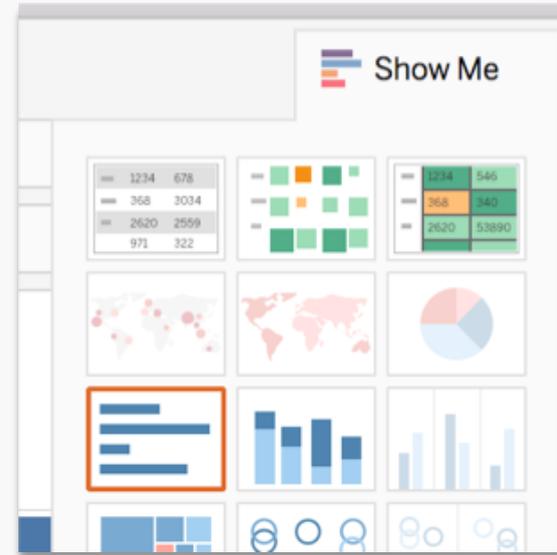
- “Learning”
- More data = things can improve.
- The more we can learn, the better for everyone.

[T]he computer is the most remarkable tool that we've ever come up with. It's the equivalent of a bicycle for our minds.

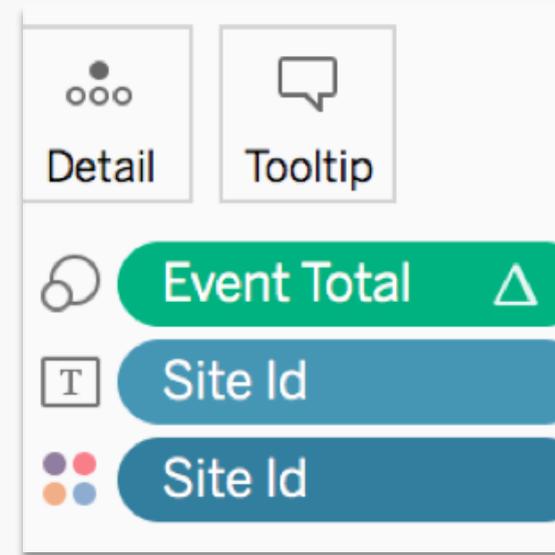
~ Steve Jobs, *Memory & Imagination*
(1990)

Ways to be Smart

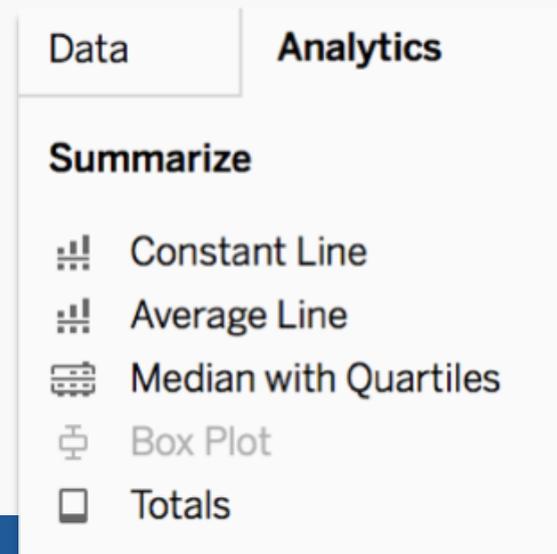
Heuristics/rules



UX



Data Science



Machine Learning

```
1 #5: 1 fetch (accuracy/accuracy/Mean:0); 2 feeds -----  
1/x_plus_b/add:0:DebugIdentity[0]":  
2  
7678e+00, -1.10659754e+00, -1.30480409e+00,  
5887e+00, nan, -1.09912372e+00,  
nan, 5.04049063e-01, nan,  
9215e-01, -7.11282551e-01, nan,  
5129e+00, -2.44008899e+00, -3.18187118e+00,  
nan, nan, nan,  
nan, -2.55820012e+00, nan,  
nan, nan, 1.91354287e+00,  
nan, -1.95995462e+00, nan, -1.15982734e+00,  
nan, -1.70120013e+00, nan,  
2090e+00, nan, -1.76648915e+00, -1.96944058e+00,  
nan, nan, -2.87428522e+00, -2.71930361e+00,  
nan, nan, nan, -2.00154305e+00,  
nan, nan, nan,  
nan, nan, nan, -2.00371289e+00,  
nan, nan, nan, -3.08434916e+00,  
0751e-01, nan, -2.70202112e+00, -9.03439403e+00,  
nan, nan, -1.23527932e+00,  
5020e+00, nan, nan, -2.32577586e+00,  
5275e+00, -2.29148245e+00, nan,  
4347e+00, -1.60913956e+00, nan, -3.40766573e+00,  
nan, nan, -2.24779010e+00,  
n): 0.00% -[84]  
1/x_plus_b/add:0[0]
```

Machine Learning: a very very short intro

Data
(+ Labels)

training

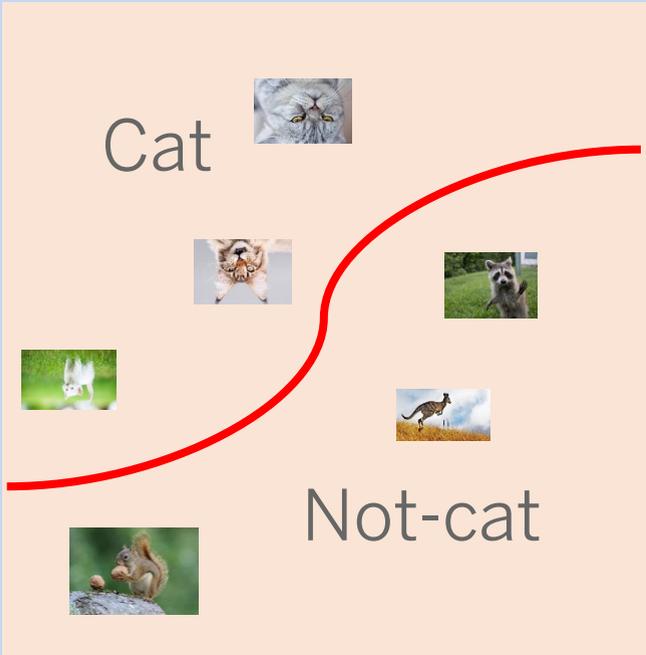
Model

generalization

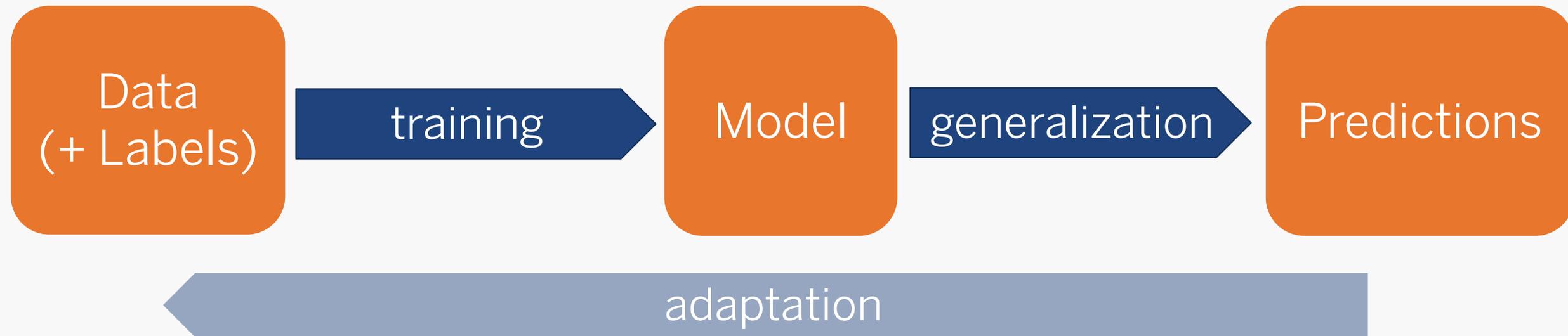
Predictions

Cat

Not-cat

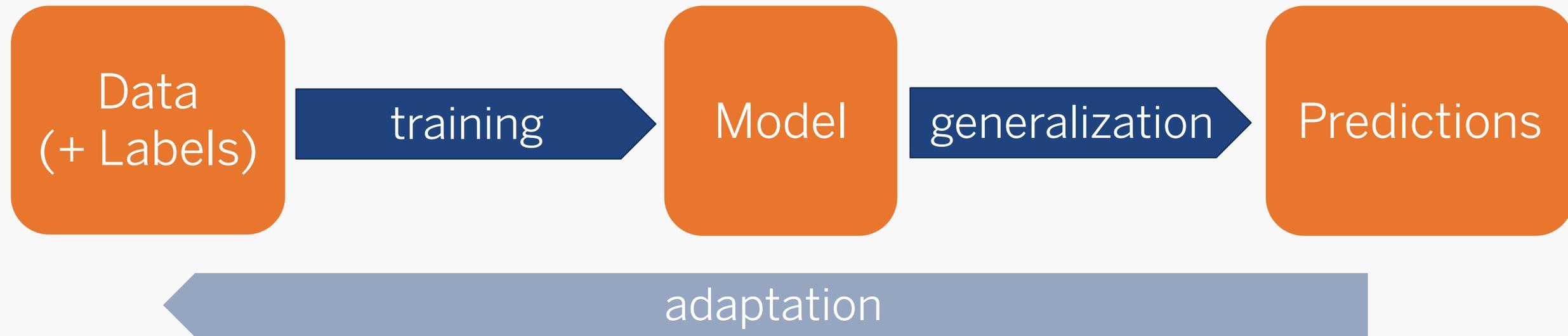


Machine Learning Key Points



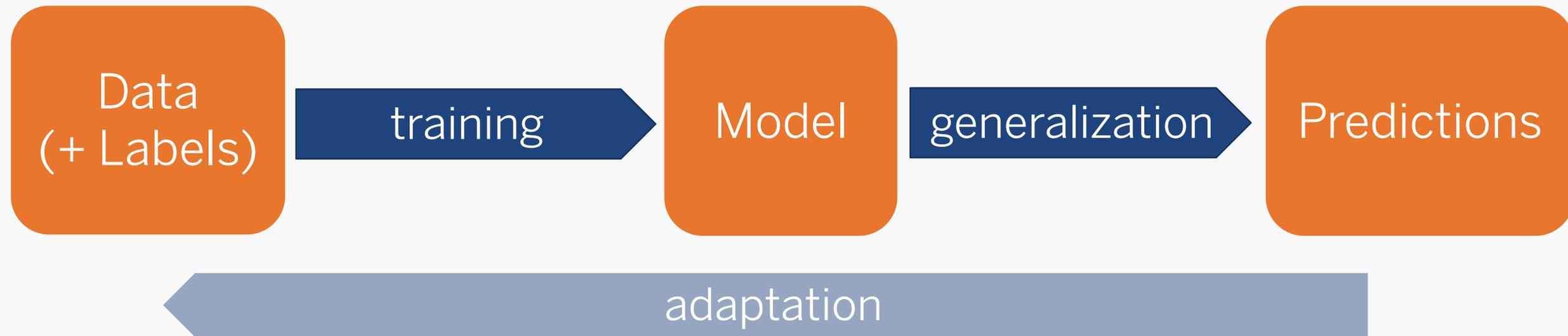
Data-Driven
Generalization
Adaptive

Machine Learning Key Points



Being data-driven, generalizable and adaptive doesn't just let us do things **better**, it enables us to do things that were previously **impossible** (or at least impractical).

Machine Learning Key Points



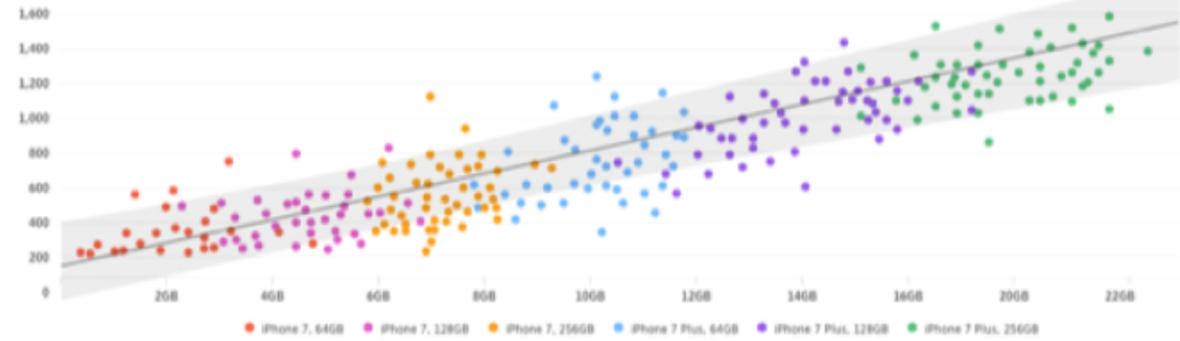
BUT, it's not going to solve all our problems

- Requires formally-specified problems
- Requires “good” data, and plenty of it

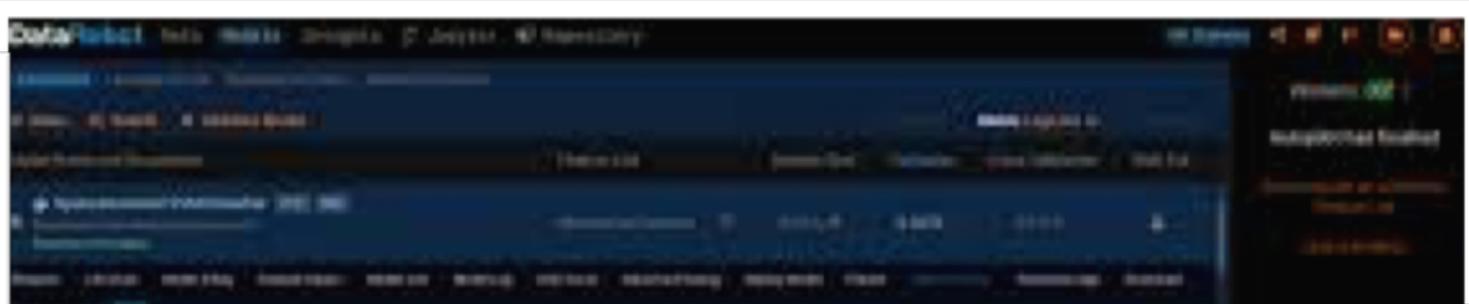
How to get from data + questions to features?

For income bracket \$250,000+ in Orange County, CA, there is a strong positive correlation between iPhone 7 units sold and data utilization with **correlation coefficient of 0.72**.

Total Units Sold vs Total Data Utilization



Was this insight useful?



SEARCH PINBOARDS MANAGE DATA ADMIN

Revenue

SUGGESTIONS

- by customer segment
- by marketing campaign
- and ROI by quarter
- and ROI by customer segment
- by product category
- and margin by product category
- count issues with growth queries on chasm worksheet

Your Recent S

number of mpls

rep = michelle wat

james lin

Revenue

sort asc

Maximize Daily Quantity in the Customer-Data-Q416 Dataset

DailyQuantity by Store when Promotion is Display

Recommended insights on: What Happened

Relating to: Promotion: Display X Store: San Francisco X

Results

- Store San Francisco and Promotion Display went up
- Store Boston and Promotion is Coupon went up
- Week is April 1-7 when Store is Miami and Promotion is Display
- Week is April 1-7 when Store is Dallas and Promotion is Display

San Francisco when Display

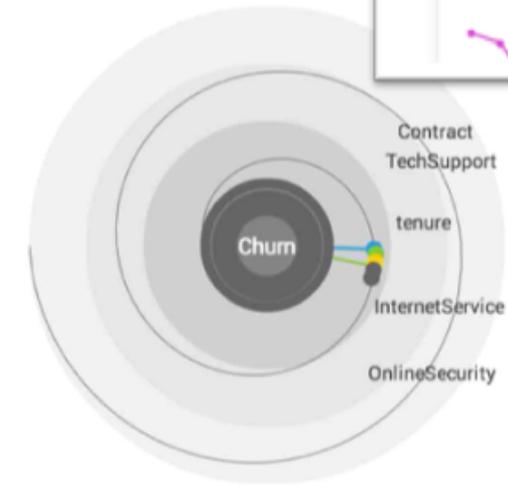
331	1,530	506,430
Count	Average	Total
Standard Deviation:	1,036	
Difference from Overall:	1,195	

You should also be aware of the case when Promotion is Display, because DailyQuantity behaved differently here. If you compare the two bars, you will see the following interesting behavior:

- Here are some cases where DailyQuantity was better than average
 - San Francisco is 5833.10 above average
 - Boston is 5481.30 above average
 - Little Rock is 5206.80 above average
- Here are some cases where profit was worse than average
 - Detroit is 5231.10 below average
 - New York is 5203.30 below average

ANALYSIS DETAIL

21 input fields were fields were potential



What else is interesting about these fields? [View All](#)

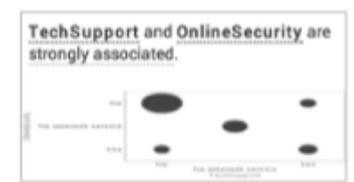
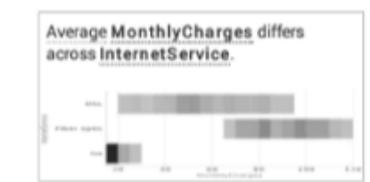


Tableau Recommendations

The screenshot displays the Tableau interface with two data sources: 'Sales' (Microsoft SQL Server) and '_workbooks' (PostgreSQL). The 'Sales' data source is selected, and a join dialog is open, showing an inner join between 'Account' and 'Opportunity' tables. The dialog includes options for Inner, Left, Right, and Full Outer joins, and a table with columns 'Data Source' and 'Opportunity' showing the join condition 'Id = Portal_Account_Id...'. Below the dialog, a table of recommended tables is visible, including 'Account_Informati...', 'Account_Informati...', 'Account_Maintena...', and 'Account_Type_Det...'. The right side of the interface shows the '_workbooks' data source with a table of recommended tables, including '_users', '_views_stats', '_projects', '_views', 'New Custom SQL', and 'New Union'. The bottom right corner shows a table with columns '#', 'Name', and 'Workbook Url' containing rows for '_workbooks'.

Connections Add

Sales
Microsoft SQL Server

Database
ALPO

Table
All 389 **Recommended 5**

These tables are popular at your organization.

- ProductMaster
- Territory
- UnifiedOwners
- AccountMaster
- AccountOwnerActivity
- New Custom SQL
- New Union

Account + Opportunity

Connection
 Live Extract

Account — Join — Opportunity

Join

Inner Left Right Full Outer

Data Source		Opportunity
Id	=	Portal_Account_Id...
Add new join clause		

Sort

Abc	Abc	Abc	Abc
Account	Account	Account	Account
Account_Informati...	Account_Informati...	Account_Maintena...	Account_Type_Det...

Connections Add

PostgreSQL

Database
workgroup

Table
All 164 **Recommended 4**

Table
These tables are popular at your organization.

- _users
- _views_stats
- _projects
- _views
- New Custom SQL
- New Union

Sort fields Data source order

#	Abc	Abc
_workbooks	_workbooks	_workbooks
Id	Name	Workbook Url

Recommendations

- What did we learn from working on recommendations?
- How did this help us think about ML? What decisions did we make?





Connect

To a File

- Microsoft Excel
- Text file
- JSON file
- PDF file
- Spatial file
- Statistical file
- More...

To a Server

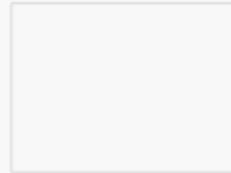
- Tableau Server
- Oracle
- Amazon Redshift
- Google BigQuery
- Microsoft SQL Server
- More...

Saved Data Sources

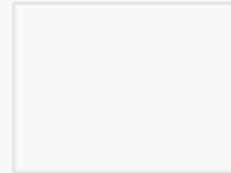
- Sample - Superstore
- World Indicators

Open

Open a Workbook



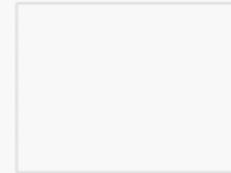
Book3 [rec-e2e-...]



Book3



Book2



Book1

Sample Workbooks

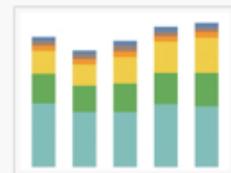
More Samples



Superstore



Regional



World Indicators

Discover

▶ Training

Getting Started

Connecting to Data

Visual Analytics

Understanding Tableau

More training videos...

📁 Resources

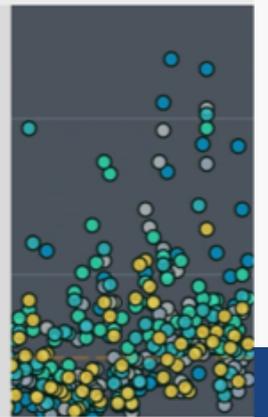
Blog - Nominate the new class of Tableau Zen Masters!

Tableau Conference - Register Now

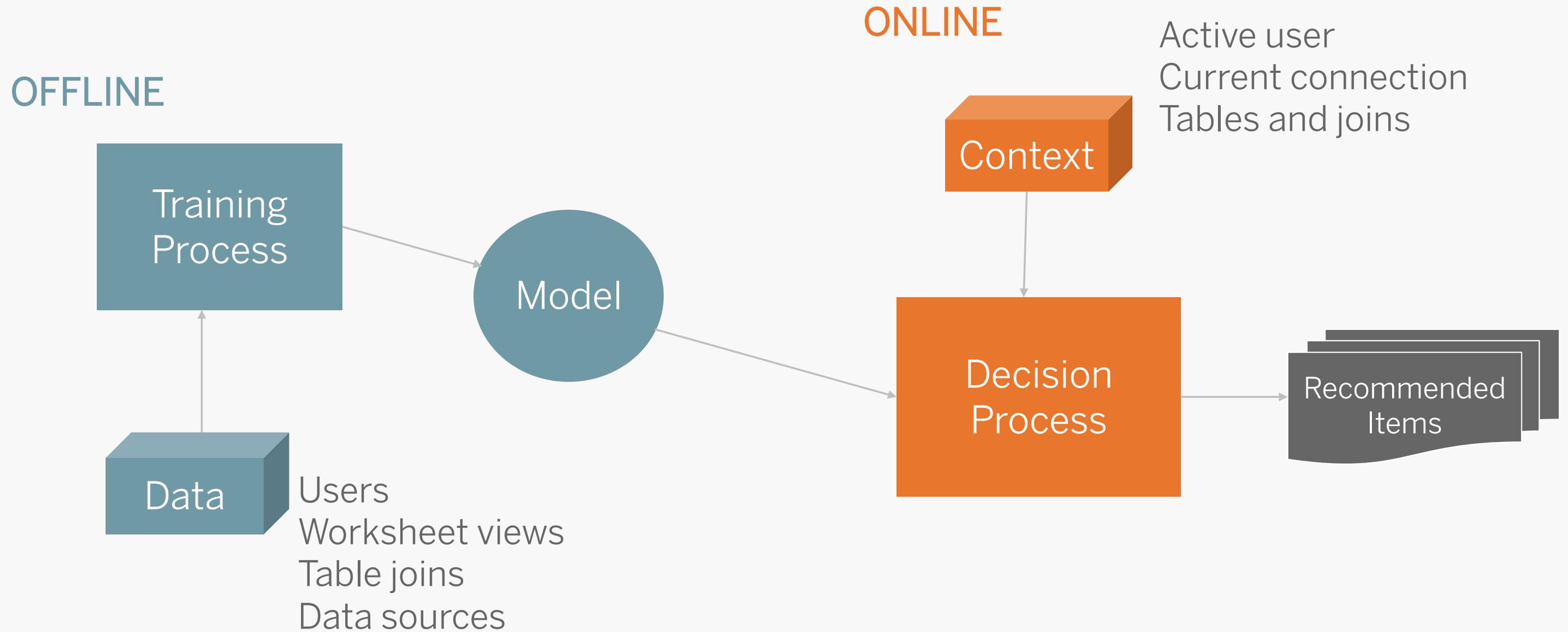
Forums

VIZ OF THE WEEK

2017: Year of the Alt-Coin →



Recommendations



Abstract Machine Learning system

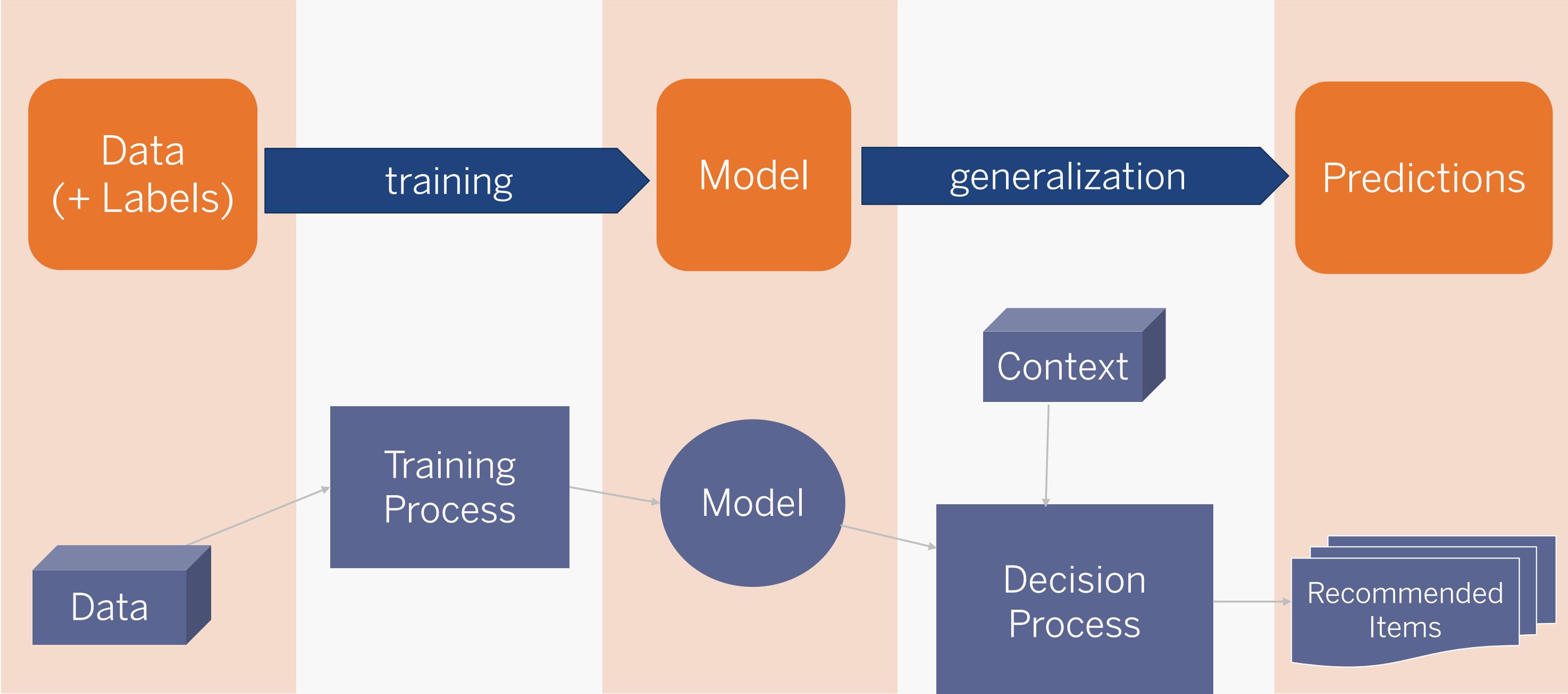


Tableau Recommendations system



Abstract Machine Learning system

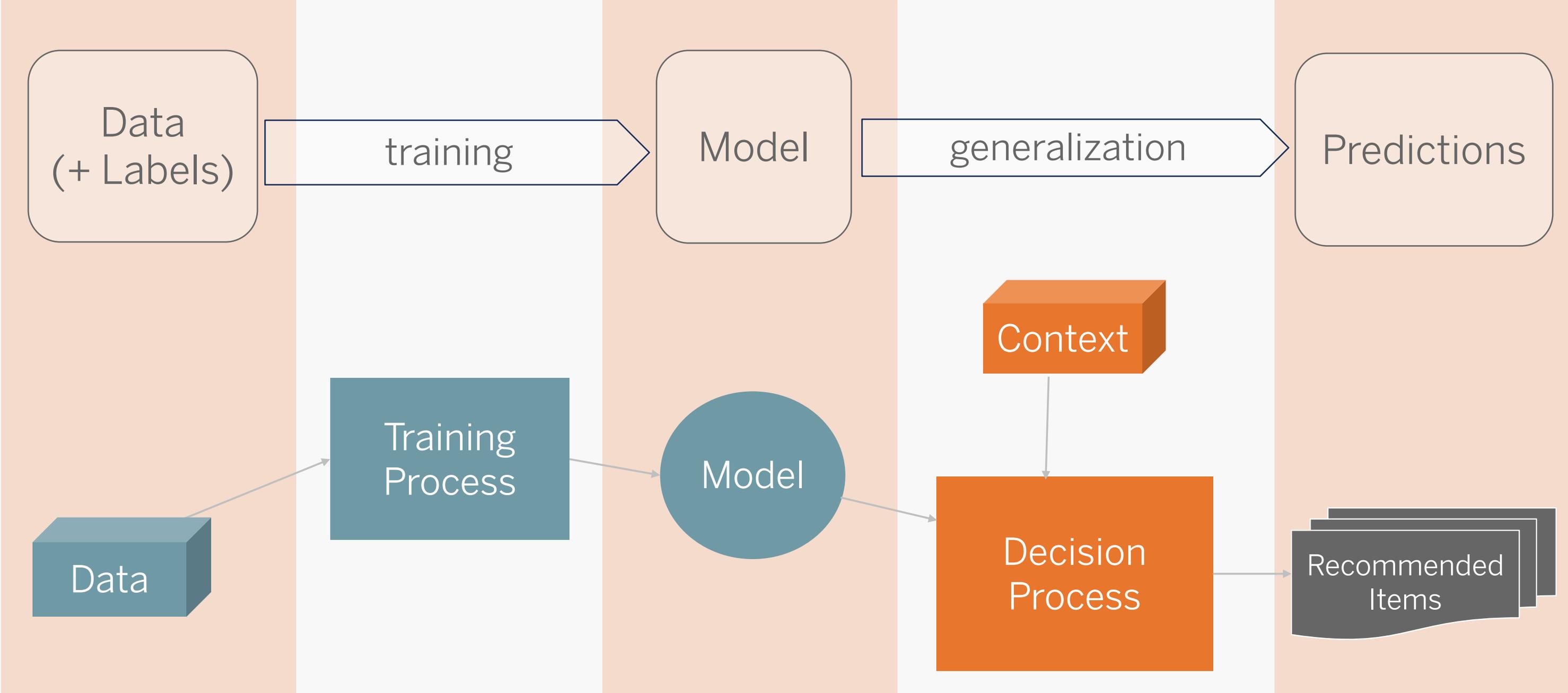


Tableau Recommendations system



Table & Join Recommendations

- Surface popular tables and joins
- Training data = table join data
- Model = associations
- Prediction = popularity

The screenshot displays a data tool interface with the following components:

- Connections:** A list of database connections, including 'alpo-dev.db.tsi.lan Microsoft SQL Server'.
- Database:** A dropdown menu set to 'ALPO'.
- Table Recommendations:** A section titled 'All 415 Recommended' with a sub-header 'These tables are popular at your organization.' It lists several tables: 'ALPO_Prod...portunity', 'ALPO_Sal...ryMaster', 'ALPO_PartnerMaster', 'Contact', and 'Technical_Sale__c'. This section is highlighted with an orange box and labeled 'recommendations'.
- Stored Procedures:** A list of stored procedures, including 'ALPO_Create...riberTables', 'sp_MSdel_cmsdimAccount', 'sp_MSdel_cmsdimCase', 'sp_MSdel_cmsdimStaff', 'sp_MSdel_c...eadContact', 'sp_MSdel_c...Assignment', 'sp_MSdel_c...aseSummary', and 'sp_MSdel_cm...Transaction'.
- Join Diagram:** A diagram showing two overlapping circles labeled 'User' and 'Opportunity', connected by a line. This diagram is highlighted with an orange box and labeled 'context'.
- Table Grid:** A table with columns containing 'Opportunity Account_Instructio...', 'Opportunity Account_Priority__c', and 'Opportunity AccountId (Opport...'. The table is sorted by 'Data source order'.
- Bottom Bar:** A navigation bar with 'Data Source' and 'Sheet 1' tabs, and a 'Go to Worksheet' button.

Data Source Recommendations

- Model = “People like you tend to like this.”
- Training data = user-workbook interactions
- Prediction = how likely user is to interact with item

recommendations

Recommended Data Sources 6

Data sources connecting to ALPO

bles here

ALPO_EmployeeAccount (ALPO) [View on server](#)

 Isaac Obezo  8  0

This is the IAMs Database with the Employee Information (from the domain)

Project: Default
Last extract: Live Connection

Popular Fields

- Region
- Office Name
- Full Name
- Department
- User Principal Name

[Add Data Source](#)

Case Data - Single Version of Truth (alp... [View on server](#)

 Geoffrey Nelson  4  0

Project: Product Troubleshooting - Repro
Last extract: Live Connection

Popular Fields

- CreatedDate
- CaseNumber
- User_Name
- Resolved Support Team

personalization

Eric Brochu

6. Table join recommendations from foreign keys and database views

Xiaojian Wang

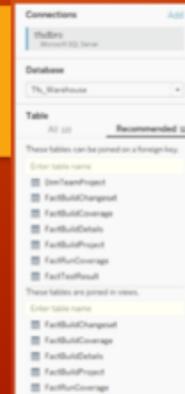
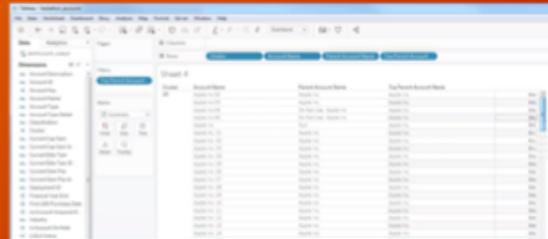
KDA: Key drivers/influencers analysis (Live demo)

Rec Team

1. Common data prep problem: string deduplication

Walther Maciel

- When data is typed in manually, especially by multiple individuals, it is bound to have inconsistencies in spelling and/or punctuation. This problem can waste many hours trying to coalesce all the different versions to a single one
- The approach that provided reasonable results in a timely manner was the token collision with double metaphone
- In a Nutshell: every string is tokenized by the double metaphone algorithm, which makes similar sounding strings

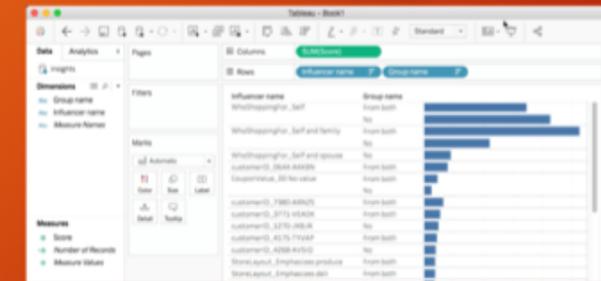


What is key driver analysis?

- KDA gives you the relative importance of an independent (driver) field as a predictor of a dependent (target) field. KDA looks for statistically significant linear relationships b/n drivers and targets. For continuous target variables, multivariate linear regression or variants (e.g. Kruskal) are common, for discrete target variables Chi-squared is common

What did we do?

- Suppose that the user is interested in comparing and contrasting drivers to sales in two different countries, or drivers for different marketing campaigns. In Tableau, the user may have a visualization of [Sales] selected as a driver, and interested in two countries, providing



2. Automatically Finding Joinable Datasources

David Mosimann

3. Automatic Statistician

Eric Brochu

- Automatic Statistician is a tool that uses Machine Learning techniques to build an artificial intelligence system for data science
- It analyzes data points in a set and identifies interesting attributes of the data, such as periodicity, noise, change points, and communicates it to an analyst
- There is some existing code from the Tableau Machine Learning Group, which was used in Eric's PhD thesis
- This hackthon was to extend the existing code to show how such a system could help explain time series data using Bayesian model selection, which was discussed in Eric's PhD thesis



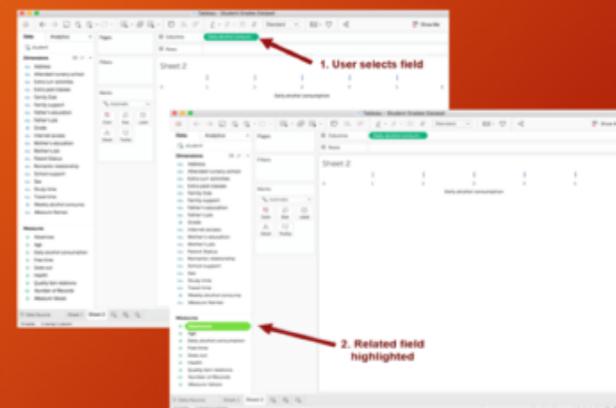
- An experiment to determine if, through data analysis, existing Tableau Server datasources can be automatically joined with the users current datasource
- Techniques using Bloom Filters were shown to be very promising
- Discussed with Anushka to determine applicability and the approach will be used in Maestro workflow

Current Datasource			Additional Datasource's	
Team	Name	Measure	Feature Team	Strategy Area
Recommendations	David	10	Recommendations	SSC
Recommendations	Ivallo	7	Content Model	SSC
Content Model	Doyle	12	Employee	Location
			David	Vancouver
			Den	Denver

5. Smart Pills

Vishaal Kapoor

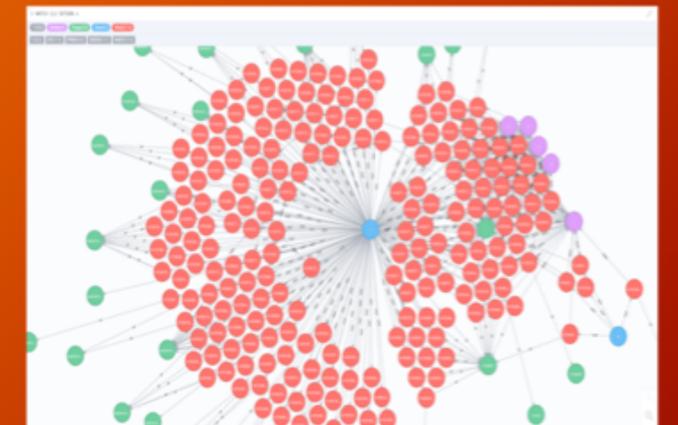
- Smart Pills highlights fields related to those that you're working with so that you can get straight to the most insightful visualizations of your data. As you drag fields to the rows and columns shelves, relevant fields that you have not yet considered are highlighted so that you know there is an interesting relationship to explore.
- In real-time, an association rule mining algorithm that is discovering statistically significant associations between the fields you're considering and those you haven't yet looked at
- The workflow has similarities to Recommended Table & Joins, but for fields instead of Tables



4. Content Model Graph Search

Mark Siegel

- Storing the full TWB/TDS content in a graph repository
 - Each node is an element (datasource, connection, field, worksheet)
 - Experimented with Titan and Neo4j
- Direct relationships between published or shared artifacts
 - Example: A direct reference from workbooks to their published datasource
 - Example: Connections to the same DB are shared
- Content searches in this model can specify exactly where they want to search
 - Example: find as a field vs find in title
- Content searches in this model can depict connections
 - Example: Ensure that all terms are descendants of a single worksheet





Open Discussion with Panel

Discussion #1:

When it comes to machine learning for my organization, where do we start? How do I know it is the right investment? What should my team be focused on?



Discussion #2:

How do you see natural language processing (NLP) playing into machine learning?

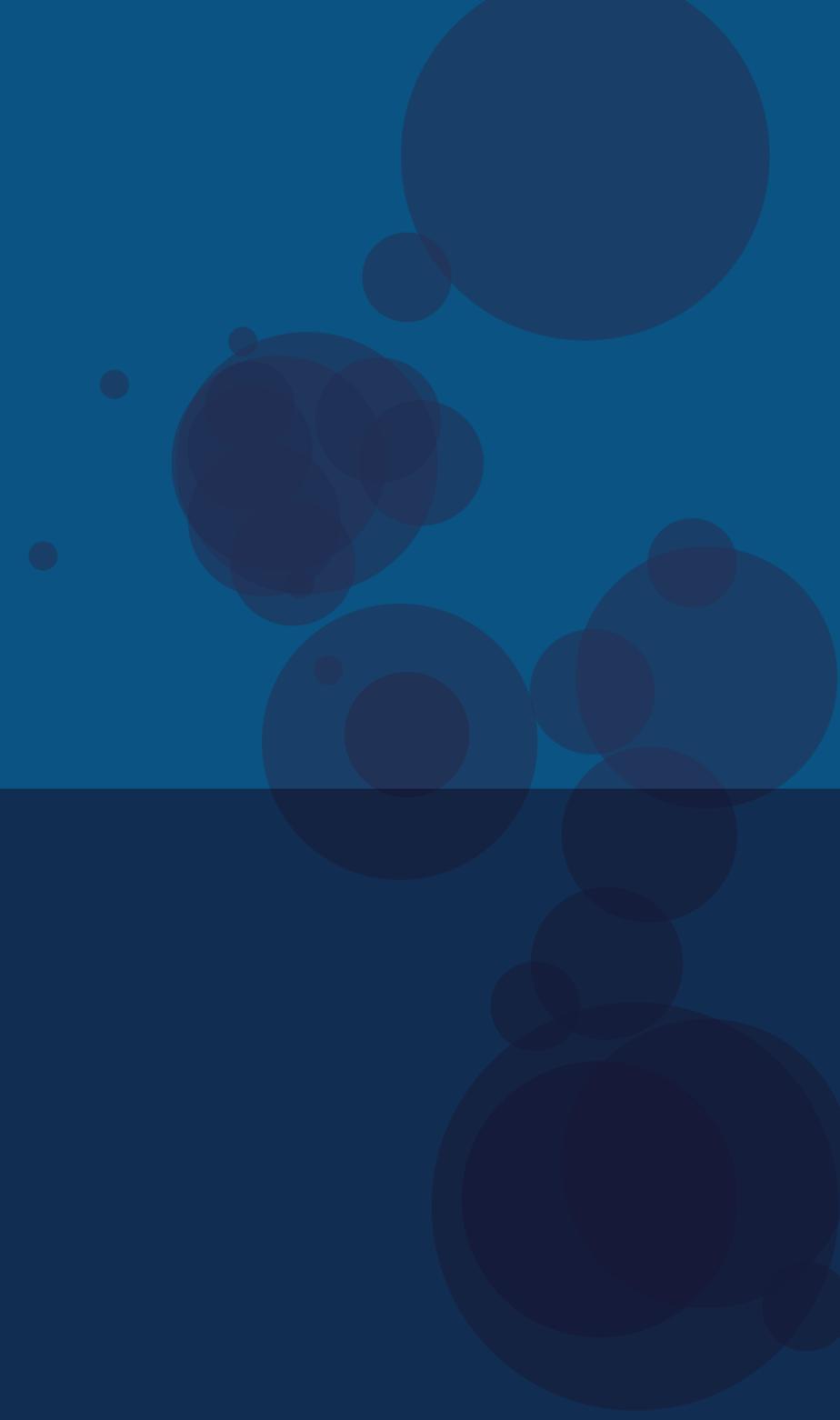


Discussion #3:

What gets you most excited when we can combine the power of machine learning and data analytics?



Additional Resources



BI Trends Additional Resources

2018 BI Trends – Full Report

<http://tabsoft.co/2A2Mwhm>



1 Don't Fear AI

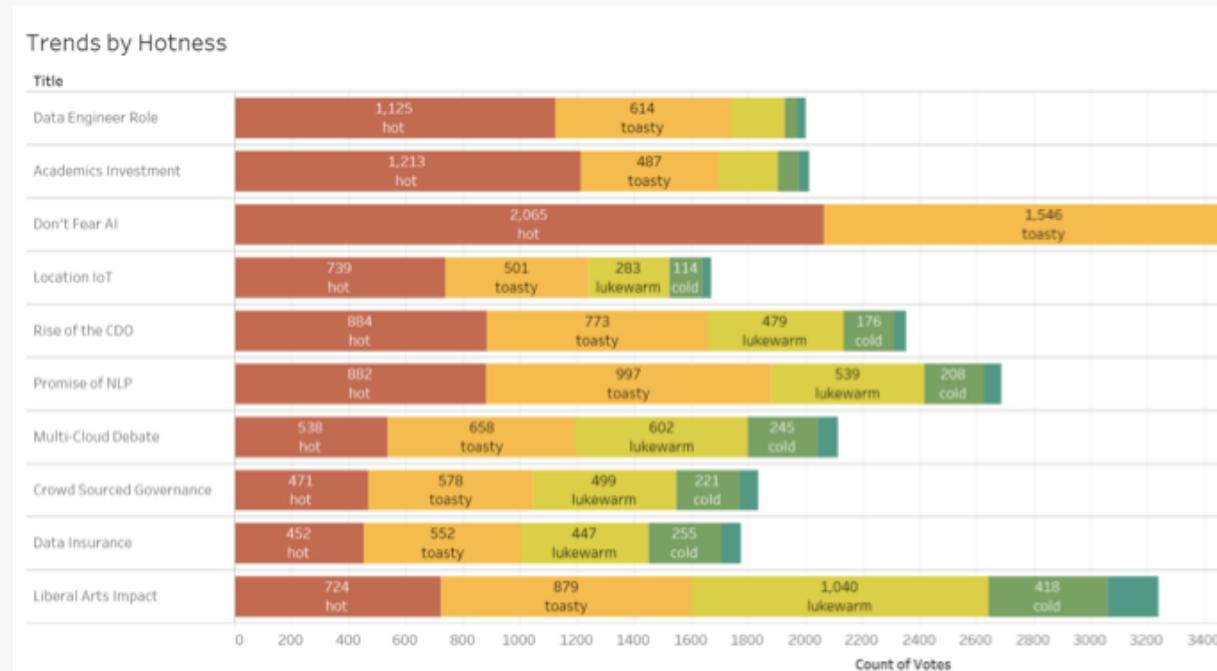
How Machine Learning Will Enhance the Analyst

Popular culture is fueling a dystopian view of what machine learning can do. But while research and technology continue to improve, machine learning is rapidly becoming a valuable supplement for the analyst. In fact, machine learning is the ultimate assistant to the analyst.

Imagine needing to quickly look at the impact of a price change on a given product. To do this, you would run a linear regression on your data. Before Excel, R or Tableau, you had to do this all manually and the process took hours. Thanks to machine learning, you can now see the product's consumption in a matter of minutes, if not seconds. As an analyst, you don't need to do that heavy lifting, and you can move onto the next question—were the higher consumption months due to an extrinsic factor such as a holiday? Was there a new release? Was there news coverage influencing product purchase or awareness? What you're not thinking about is how you wish you could have spent more time perfecting your regression model.

BI Trends Voting Visualization

<https://alpo/authoring/2018Top10BITrendsVotes/Votes?#1>



BI Trends Additional Resources

2018 BI Trends – Facebook Live w/ Chief Product Officer Francois Ajenstat

<http://tabsoft.co/2D653LE>



We're chatting with Tableau Chief Product Officer Francois Ajenstat,...

16K views · November 16, 2017

BI Trends “11th Trend” – Facebook Live

<http://tabsoft.co/2mlkBzV>



The Tableau community has spoken! Find out what they chose...

12K views · December 13, 2017



Thank you