Microsoft | Innovation Center Torino

CloudTech

Marrakesh 2016

Microsoft | Innovation Center Torino

CloudTech

Marrakesh 2016

Introducing Microsoft Azure Machine Learning

Claudio Rossi

Agenda:

Overview of Machine Learning

Overview of Azure ML

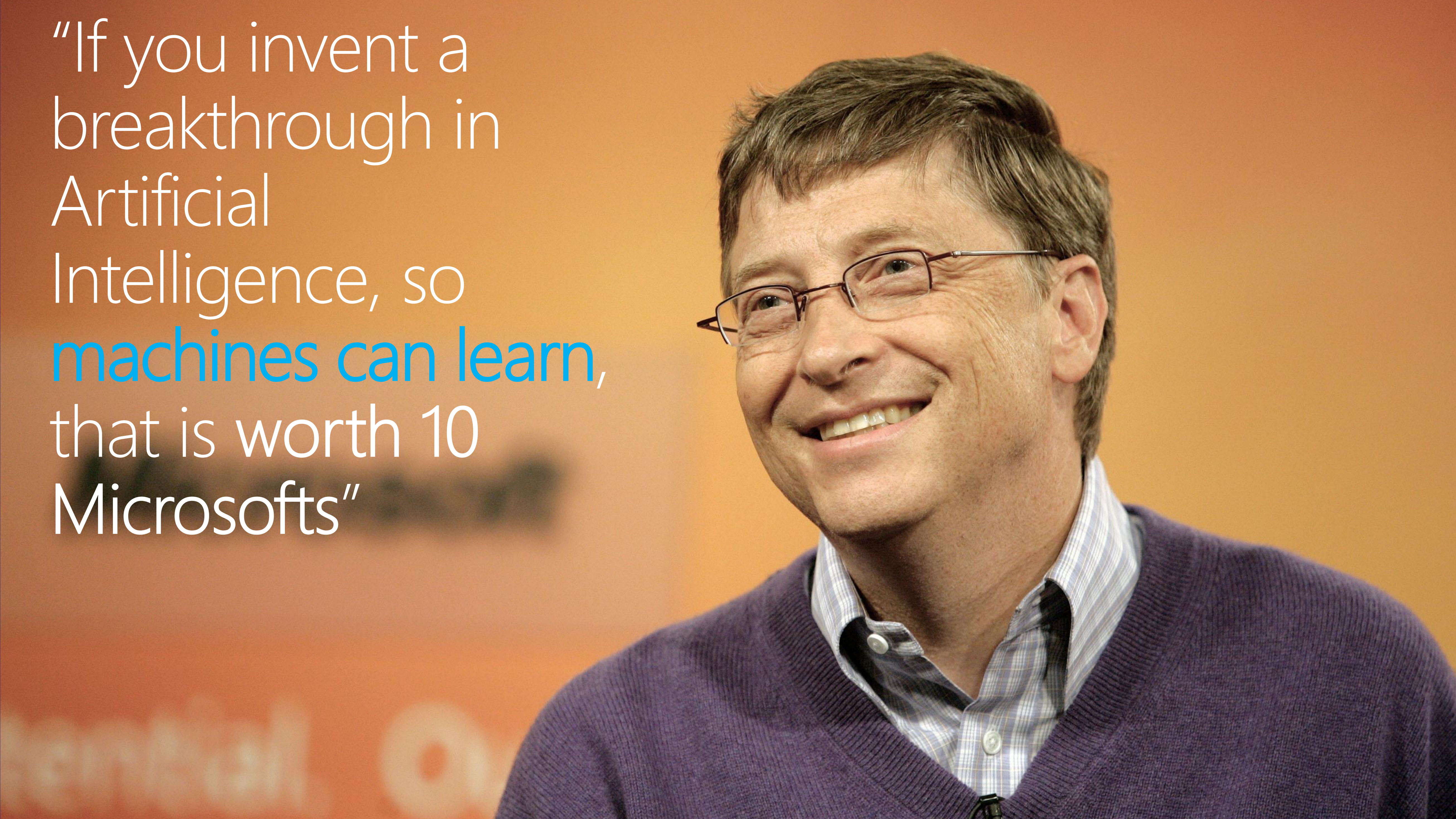
Guided Demo

Hands-on Lab

Overview of Machine Learning

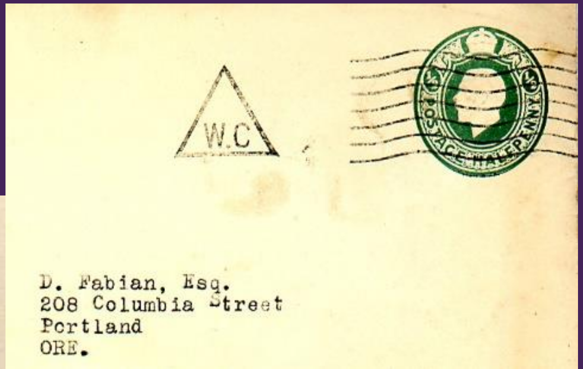
I believe over the next decade computing will become even more ubiquitous and intelligence will become ambient…This will be made possible by an ever-growing network of connected devices, incredible computing capacity from the cloud, insights from big data, and intelligence from machine learning.

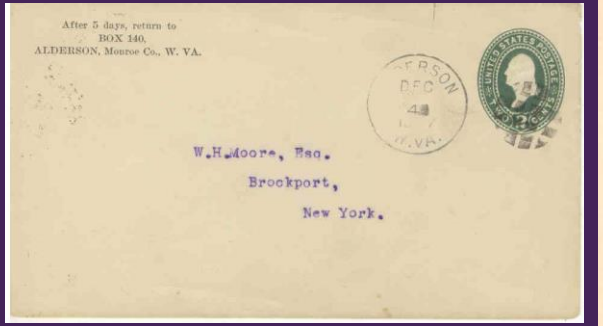
Satya Nadella
CEO @ Microsoft


"If you invent a breakthrough in Artificial Intelligence, so machines can learn, that is worth 10 Microsofts"

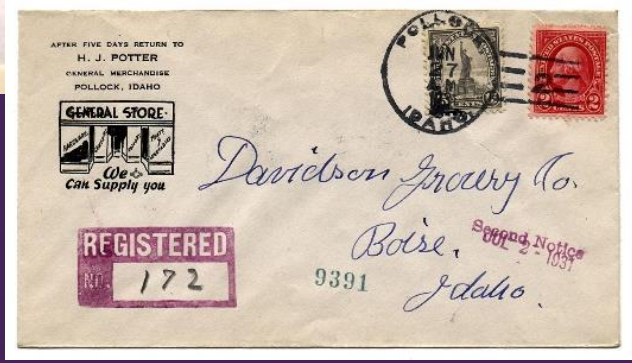

TRICHUR
Registered
ONE ANNA
Mr. Eugene Gummersbach,
15, 17, South Broadway
St Louis, Mo.
U.S.A.
NOT AT 402
42251


U.S. ARMY POSTAL SERVICE
MAY
1944
PASSED BY
ARMY EXAMINER
Mr George A Hughes
1135 Clarenden Crescent
Oakland
Calif.

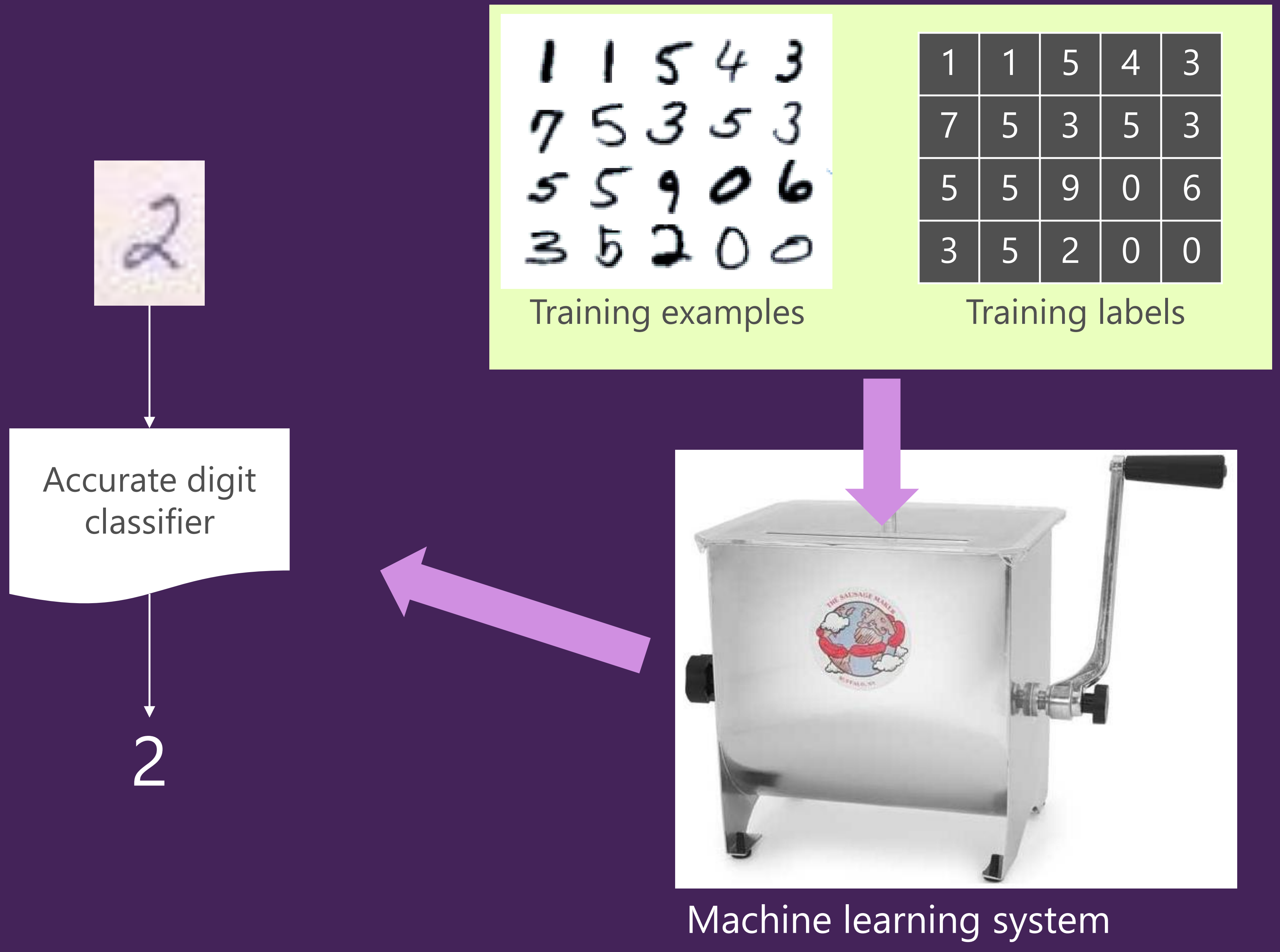

W.C
D. Fabian, Esq.
208 Columbia Street
Portland
ORE.


After 5 days, return to
BOX 140,
ALDERSON, Monroe Co., W. VA.
W.H.Moore, Esq.
Brockport,
New York.


Mrs. P. Brown
1318 - 5th Ave North
Saskatoon Sask
S2K 252
SASKATOON
I PM
17 V
1972
SASK
HELP RETAR
CHILDRE
AIDEZ L
ENFANT AP
CANADA
KAULBACH ISLAND
9c
KAULBACH ISLAND
9c
INBOUND
LOCAL CARRIAGE SERVICE
P.K.I., Ltd,
Box 220,
Chester, N.S.,
Canada


AFTER FIVE DAYS RETURN TO
H. J. POTTER
GENERAL MERCHANDISE
POLLOCK, IDAHO
GENERAL STORE
We Can Supply you
REGISTERED
No. 172
Davidson Grocery Co.
Boise,
Idaho.
9391
Second Notice


Via NYRBA Line
INSUFFICIENTLY PAID
VIA AIR MAIL
AIR MAIL
VIA NYRBA
Par Avion
By Air Mail
Mr. George S. Chapman
c/o Administrador de Correos
San Pedro de Macoris,
R.D.


Accurate digit classifier
2
1 1 5 4 3
7 5 3 5 3
5 5 9 0 6
3 5 2 0 0
Training examples
Training labels
Machine learning system

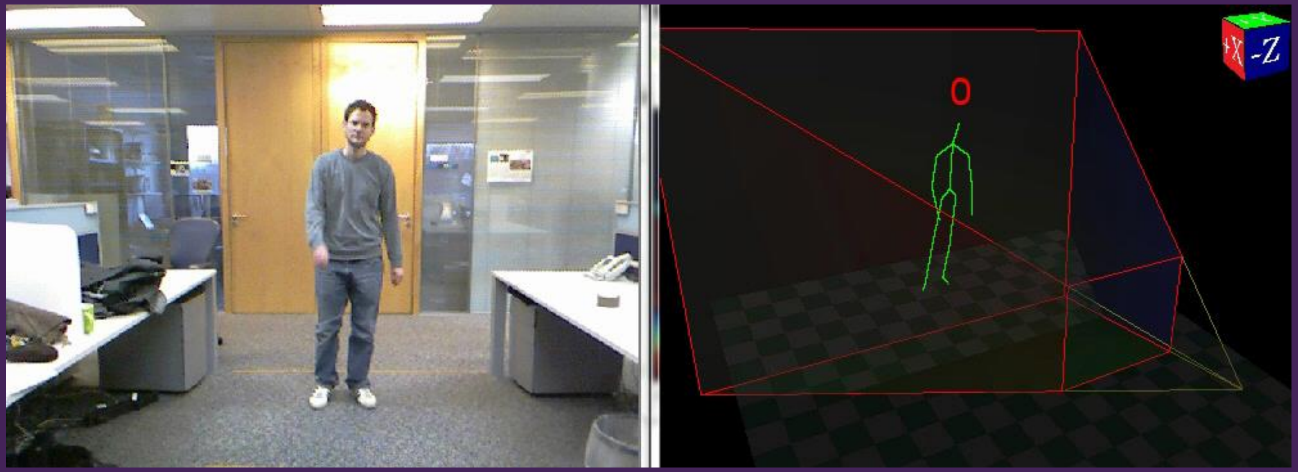






training data (expensive)

synthetic training data (cheaper)



ML system

Machine Learning (ML)

“The goal of machine learning is to build computer systems that can adapt and learn from their experience.”

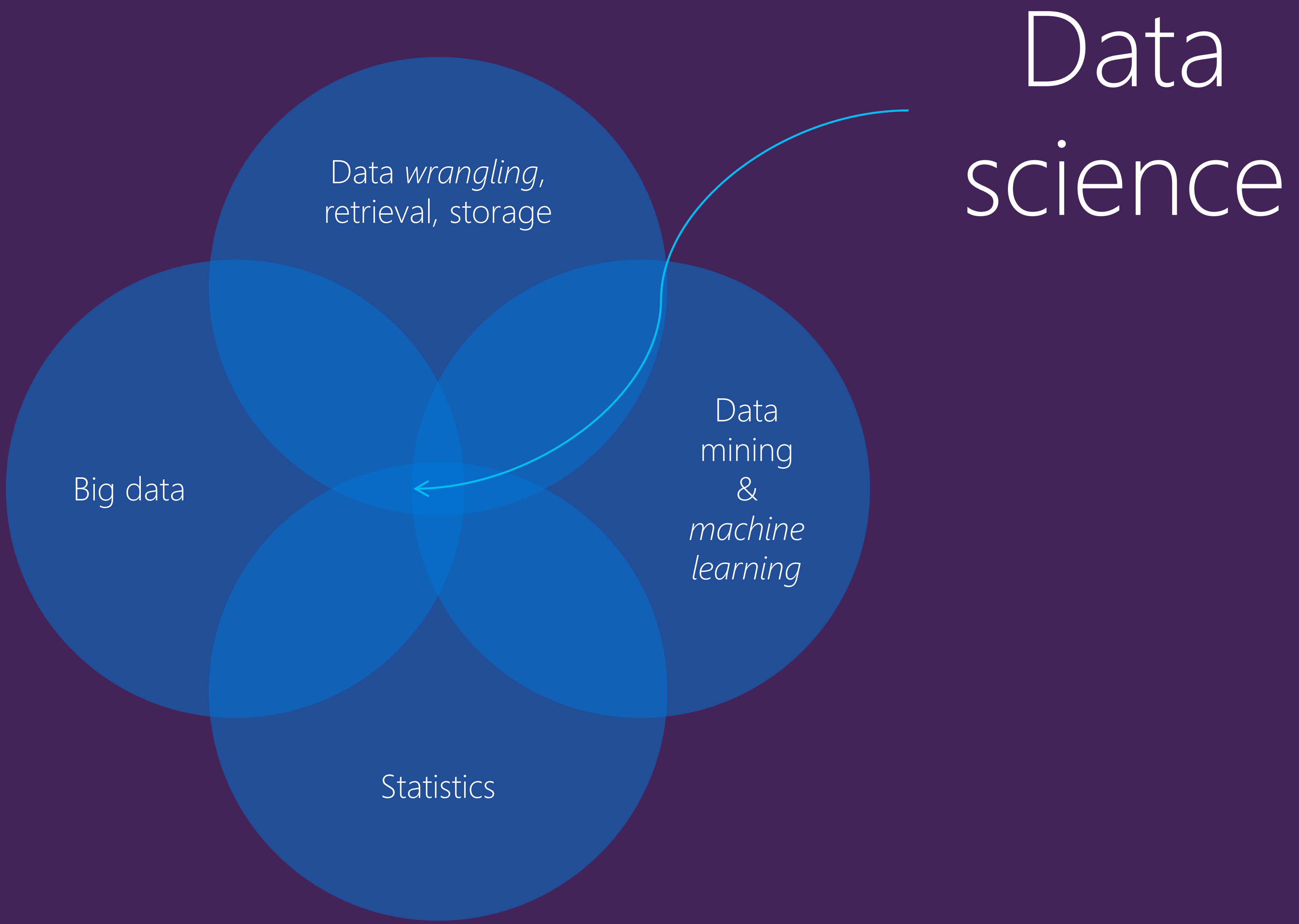
– Tom Dietterich

ML is Data Science?

Data science

- is the study of the generalizable extraction of knowledge from data *(Wikipedia)*
- is getting predictive and/or actionable insight from data *(Neil Raden)*
- involves extracting, creating, and processing data to turn it into business value. – *Vincent Granville (Developing Analytic Talent: Becoming a Data Scientist)*

Where Data Science lies?



Why learn it?

1. **Learn it when you can't code it**
   (e.g. Recognizing Speech/image/gestures)
2. **Learn it when you can't scale it**
   (e.g. Recommendations, Spam & Fraud detection)
3. **Learn it when you have to adapt/personalize**
   (e.g. Predictive typing)
4. **Learn it when you can't track it**
   (e.g. AI gaming, robot control)

ML Cycle

1. Select & initialise a model
2. Train model (process cases)
3. Validate model

   ...by scoring (making predictions) a test data set and evaluating the results

4. Use it: Explore or Deploy

   ...visualise and study

   ...deploy as a (web) service

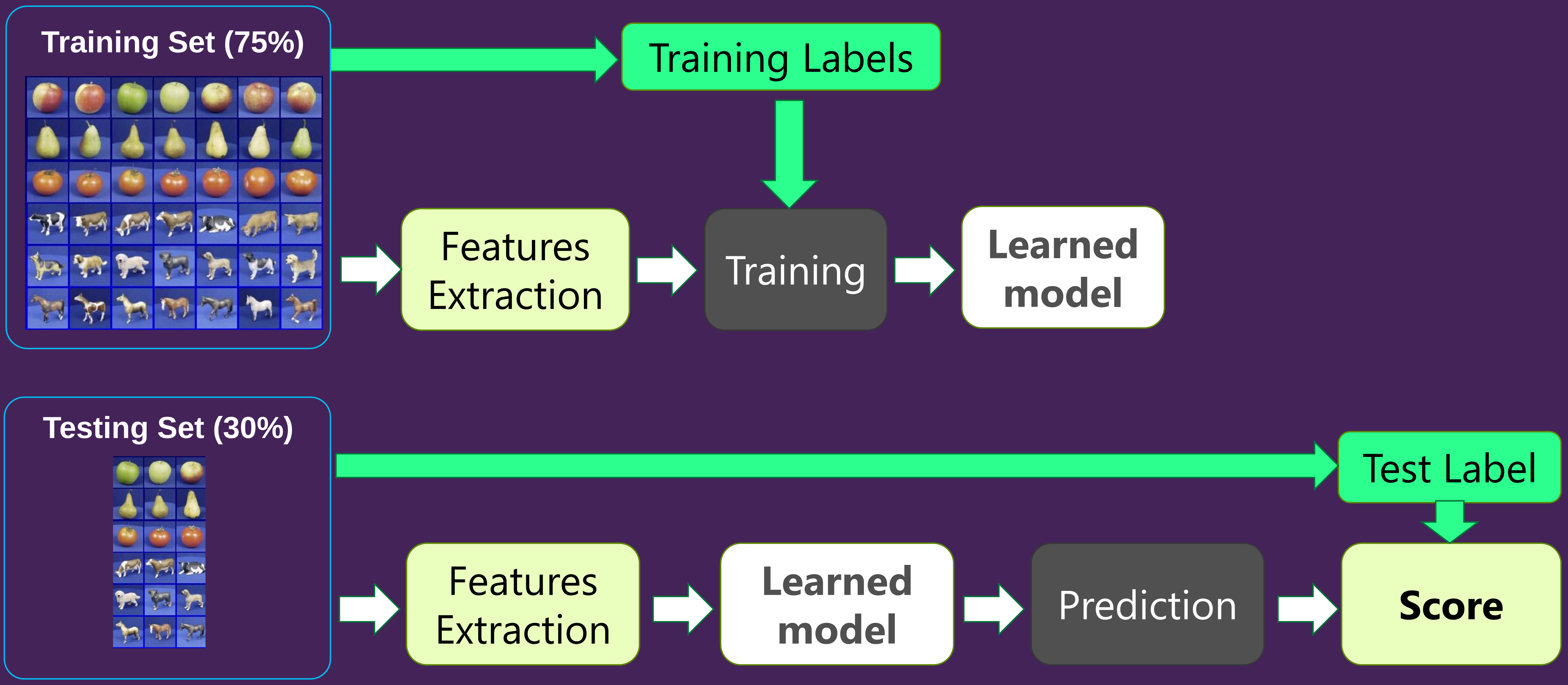
5. Update and revalidate

Algorithm Classes

Supervised	Ground truth known in the data set (regression, classifiers, …)
Un-Supervised	Ground truth not known (clustering, dimensionality reduction)

Classifiers	Assign a category to each item
Clustering	Discover natural groupings of cases
Regression	Predict numerical outcomes
Recommenders	Explore associations between cases
Ensembles	mix them up

Training & Testing a Model



Scoring a Model

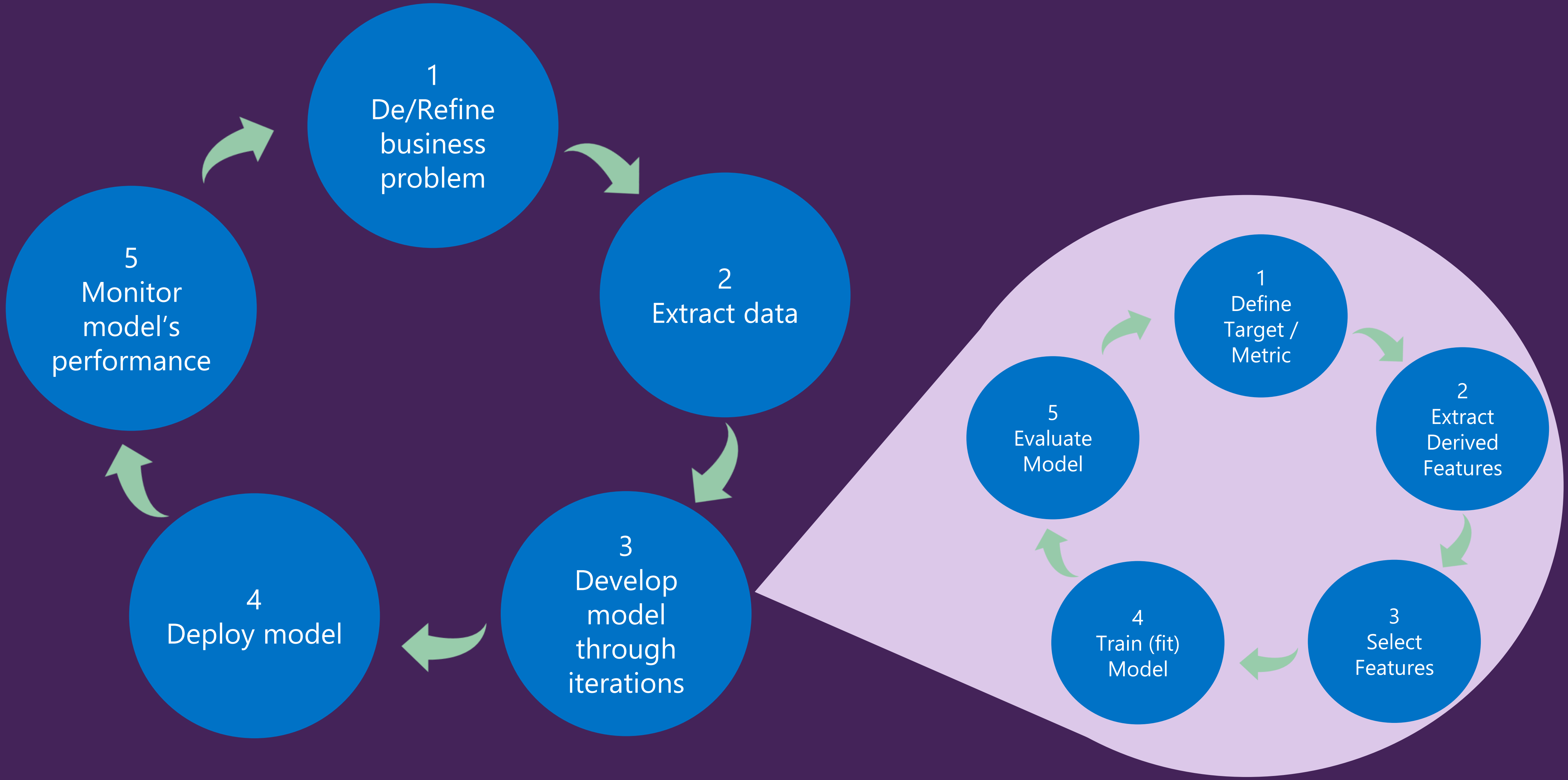
- Regression: Mean Square Error (MSE)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$

- Classification: Confusion Matrix
  - Accuracy = (TP + TN) / TOTAL
  - Precision = TP / (TP + FP)
  - Specificity = TN / RN
  - …
  - Full @ https://en.wikipedia.org/wiki/Sensitivity_and_specificity

Confusion Matrix		*Predicted* TRUE	*Predicted* FALSE
Real	**TRUE**	True Positive (TP)	False Negative (FN)
Real	**FALSE**	False Positive (FP)	True Negative (TN)

Steps to build a Machine Learning Solution



Feature engineering is the **key**…

Feature engineering: when you use your knowledge about the data to create fields that make machine learning algorithms work better.

It is easily the most important factor in determining the success of a machine learning project

How does one engineer a good feature? Rule of thumb is to try to design features where the likelihood of a certain class goes up monotonically with the value of the field.

Great things happen in machine learning when human and machine work together, **combining a person's knowledge of how to create relevant features from the data with the machine's talent for optimization**..

More data beats a cleverer algorithm…

More data wins. There's increasingly good evidence that, in a lot of problems, very simple machine learning techniques can be levered into incredibly powerful classifiers with the addition of loads of data.

Computer algorithms trying to learn models have only a relatively few tricks they can do efficiently, and many of them are not so very different. Performance differences between algorithms are typically not large.

Thus, if you want better performances:

1. **Engineer better features**
2. **Get your hands on more high-quality data**

Tools & Salaries

SQL (any): #1 data science tool

Love it, or...use it.

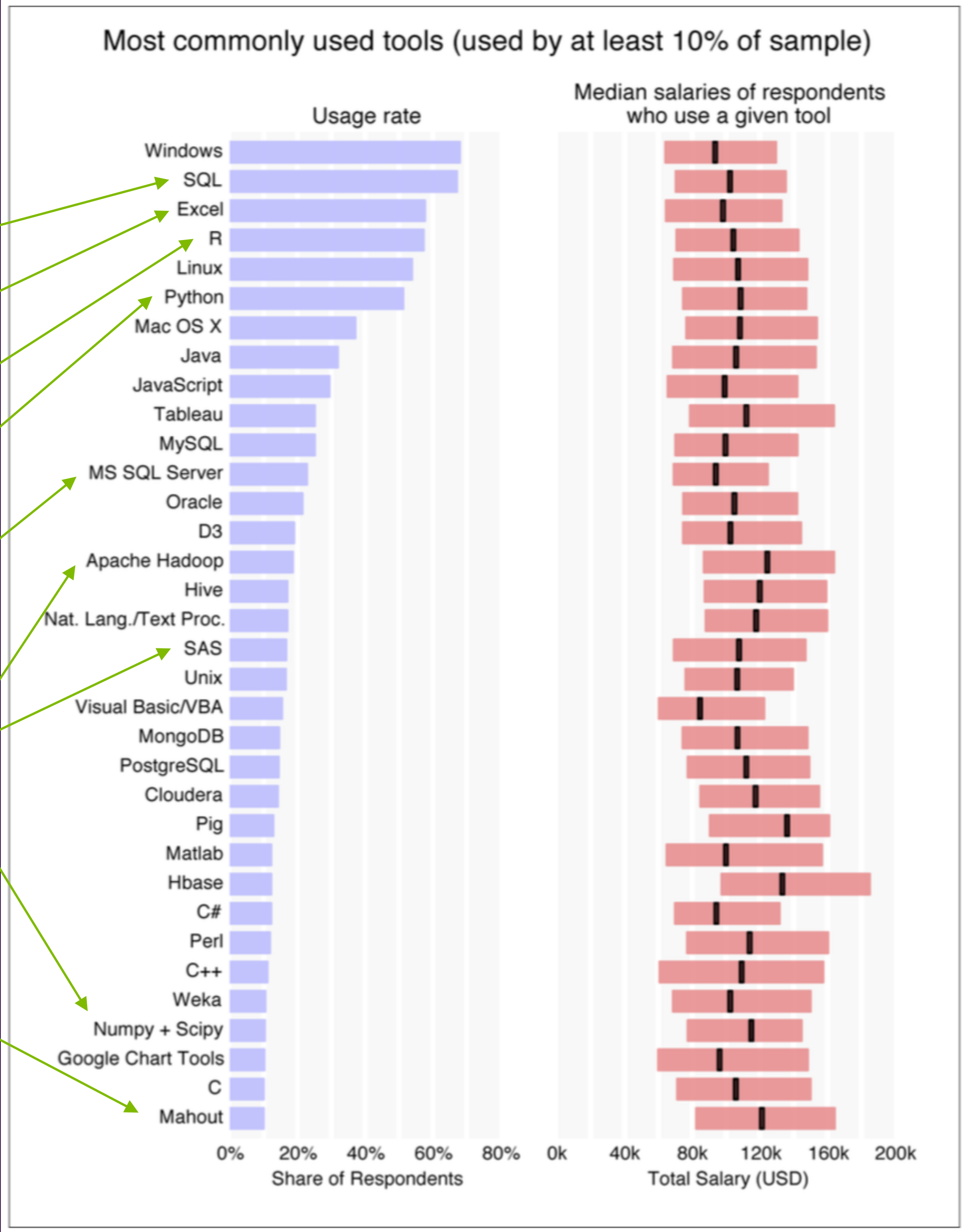
R *rulez*

Getting very important

SQL Server!

SAS: most likely BI to be discontinued by Gartner clients (Oct 2014 MQ)

Sometimes



What is R? (i)

What is R? (ii)

- Language, interpreter, *poor* IDE
- 5000+ packages of statistical software
- Better IDE: RStudio
  - http://www.rstudio.com/
  - Rattle (Data Mining GUI) makes it even easier
- Open source, free, multiplatform
  - Core R: the purest version: http://cran.r-project.org/
  - Revolution Analytics: parallelism & performance: http://www.revolutionanalytics.com/
  - Azure ML: built-in

Data Science

Complexity

Data Science is far too complex today

- Access to quality ML algorithms, cost is high.
- Must learn multiple tools to go end2end, from data acquisition, cleaning and prep, machine learning, and experimentation.
- Ability to put a model into production.

This must get simpler, it simply won't scale!

Suggested Readings

A Few Useful Things to Know about Machine Learning

Pedro Domingos
Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350, U.S.A.
pedrod@cs.washington.edu

ABSTRACT

Machine learning algorithms can figure out how to perform important tasks by generalizing from examples. This is often feasible and cost-effective where manual programming is not. As more data becomes available, more ambitious problems can be tackled. As a result, machine learning is widely used in computer science and other fields. However, developing successful machine learning applications requires a substantial amount of "black art" that is hard to find in textbooks. This article summarizes twelve key lessons that machine learning researchers and practitioners have learned. These include pitfalls to avoid, important issues to focus on, and answers to common questions.

1. INTRODUCTION

Machine learning systems automatically learn programs from data. This is often a very attractive alternative to manually

correct output y_t for future examples $\mathbf{x_t}$ (e.g., whether the spam filter correctly classifies previously unseen emails as spam or not spam).

2. LEARNING = REPRESENTATION + EVALUATION + OPTIMIZATION

Suppose you have an application that you think machine learning might be good for. The first problem facing you is the bewildering variety of learning algorithms available. Which one to use? There are literally thousands available, and hundreds more are published each year. The key to not getting lost in this huge space is to realize that it consists of combinations of just three components. The components are:

Representation. A classifier must be represented in some formal language that the computer can handle. Con-

Want to understand the ***what***, ***why***, and ***how*** of predictive analytics?

Here's a short, ordered reading list designed to get you up to speed fast…

The Signal And The Noise: Why So Many Predictions Fail — but Some Don't by Nate Silver.
Nate Silver built an innovative system for predicting baseball performance and predicted the 2008 and 2012 elections within a hair's breadth. The New York Times publishes FiveThirtyEight.com, where Silver is one of the nation's most influential political forecasters. **Why read:** This book will inspire your "predictive" imagination and ground you in realities of predictive analytics.

Predictive Analytics: The Power To Predict Who Will Click, Buy, Lie, or Die by Eric Siegel.
Eric Siegel has written a very accessible book on how predictive analytics works. It's chock-full of dozens of real-world examples, such as how Chase Bank predicted mortgage risk (before the recession), IBM Watson won *Jeopardy!*, and Hewlett-Packard predicted employee flight risk. **Why read:** *Predictive Analytics* is a perfectly paced explanation of how predictive analytics works and a repository of dozens of real examples across many use cases.

Uncontrolled: The Surprising Payoff Of of Trial-and-Error For Business, Politics, and Society by Jim Manzi.
Predictive analytics is not magic — it's science. Jim Manzi reminds us of the power of the scientific method and reveals the shocking truth that many huge societal and business decisions are made based on misinterpreted data and statistics. **Why read:** Controlled experimentation amplifies the results of predictive analytics and helps avert the risk of inaccurate predictions.

Data Mining: Practical Machine Learning Tools and Techniques by Ian H. Witten, Eibe Frank, and Mark Hall.
If you write code or have a computer science background, you'll probably want to know the gory details of how predictive analytics works. **Why read:** *Data Mining* provides a thorough grounding in machine learning concepts as well as practical advice on applying machine learning tools and techniques in real-world data mining situations.

CloudTech

Marrakesh 2016

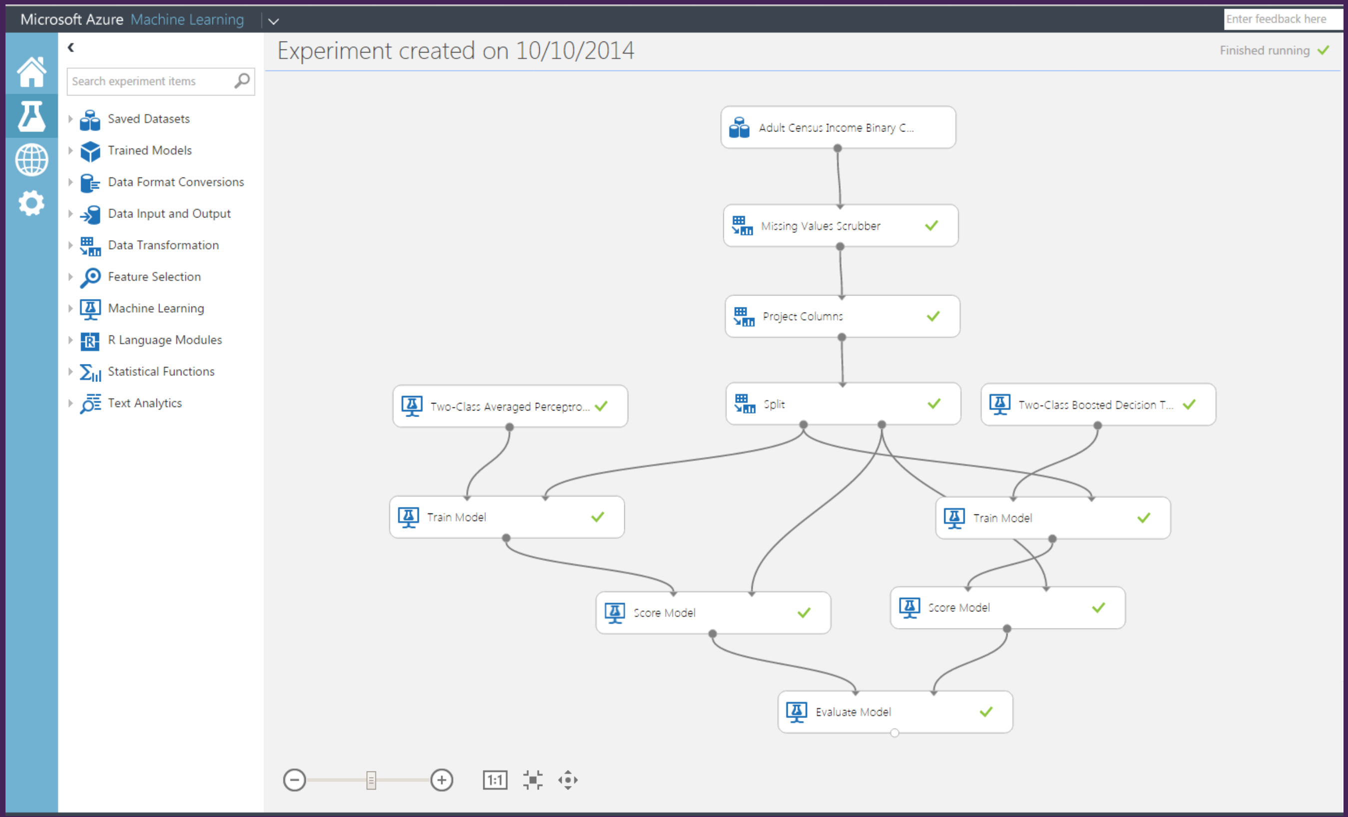
Introducing Azure ML

Microsoft Azure Machine Learning

Features and Benefits

Reduce complexity to broaden participation

- Accessible through a web browser, no software to install;
- Collaborative work with anyone, anywhere via Azure workspace
- Visual composition with end2end support for data science workflow;
- Best in class ML algorithms;
- Extensible, support for R OSS.



Microsoft Azure Machine Learning

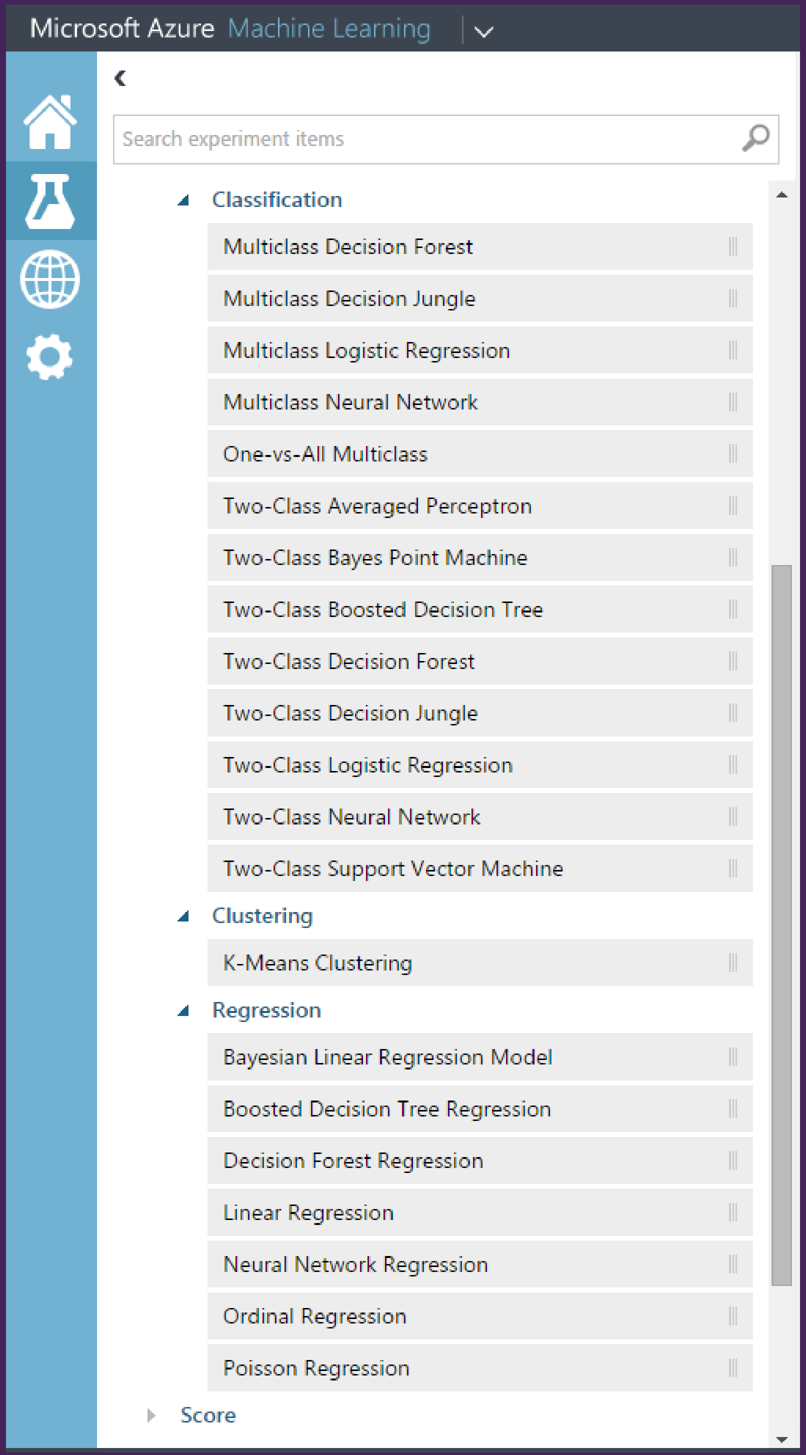
Features and Benefits

Rapid experimentation to create a better model

Immutable library of models, search discover and reuse;

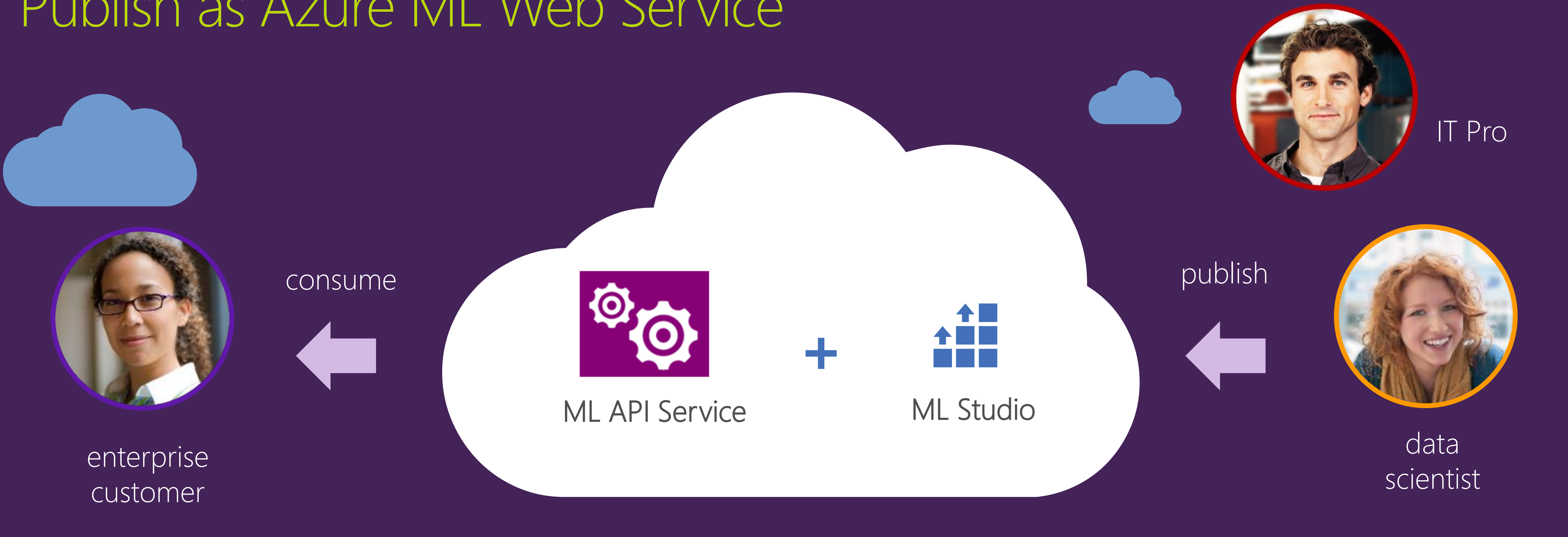
Rapidly try a range of features, ML algorithms and modeling strategies;

Quickly deploy model as Azure web service to our ML API service.



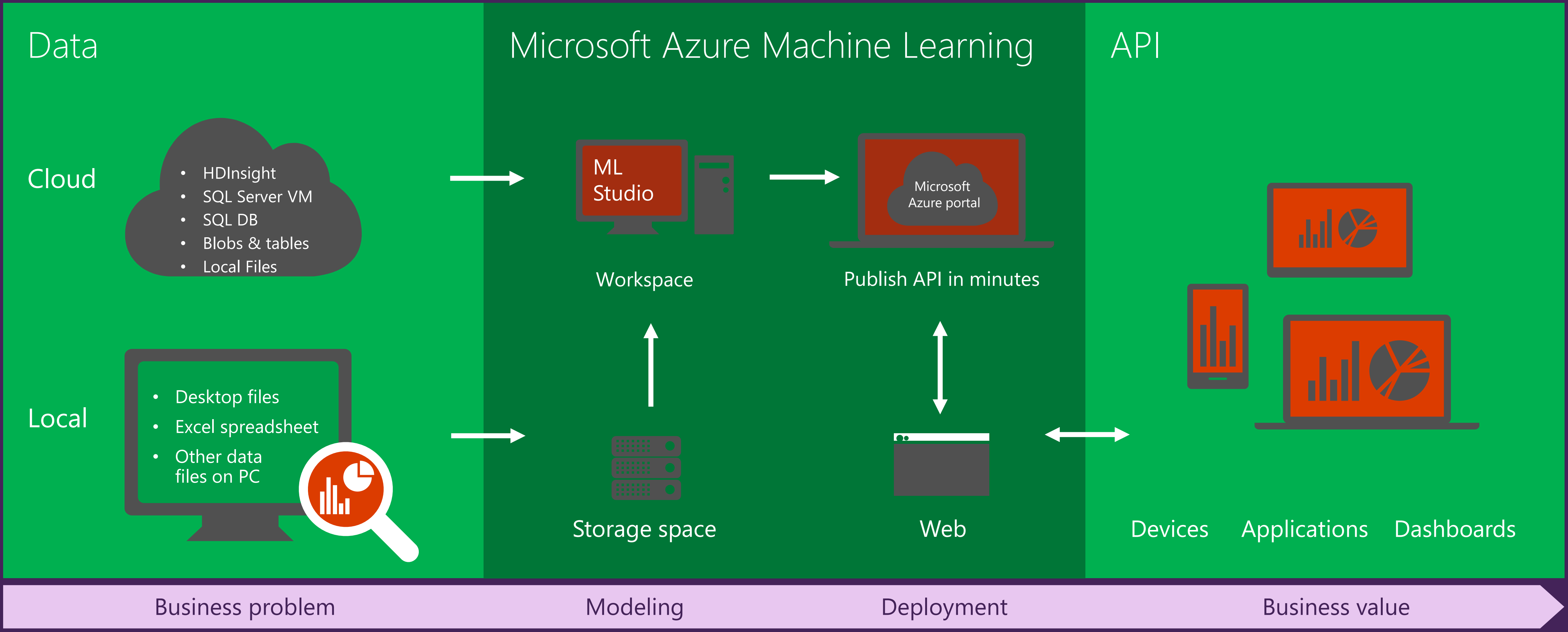
Microsoft Azure Machine Learning

Publish as Azure ML Web Service



- Highly scalable services built on Azure. Eliminate upfront costs for hardware resources.
- Scored in batch-mode or request-response mode;
- Actively monitor models in production to detect changes;
- Telemetry and model management (retrain).

Business problem to business value



Online resources

- Lots of stuff: https://github.com/Azure-Readiness
- Free SQL Server 2014 Technical Overview e-book

  microsoft.com/sqlserver and Amazon Kindle Store
- Free online training at Microsoft Virtual Academy

  microsoftvirtualacademy.com
- Try new Azure data services previews!

  Azure Machine Learning, DocumentDB, and Stream Analytics

Demo: Azure ML

CloudTech

Marrakesh 2016

Hands-On: Lab

CloudTech

Marrakesh 2016

Q&A

Presentation credits:
Raymond Laghaeian @ TechEd 2016
Rafal Lukawiecki @ Microsoft Ignite 2015

Claudio Rossi
rossi@ismb.it | @Resonance82

Microsoft | Innovation Center Torino