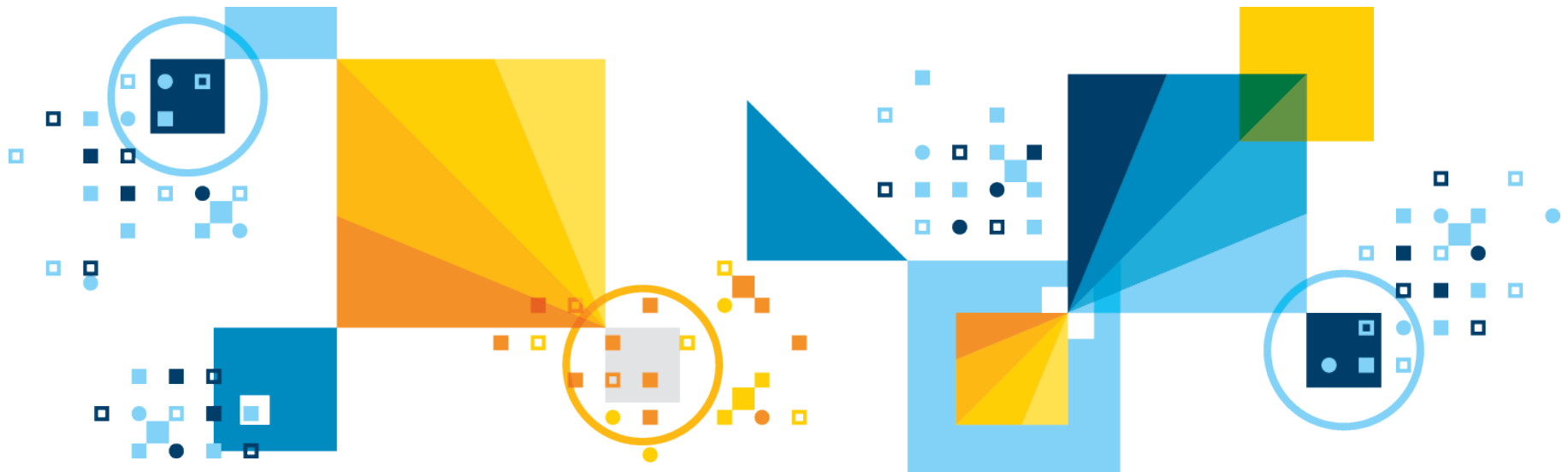


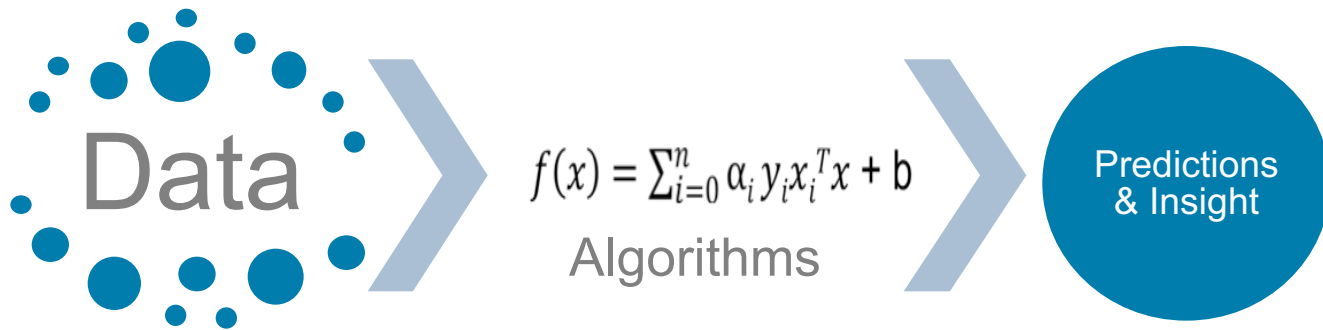
Introduction to Machine Learning



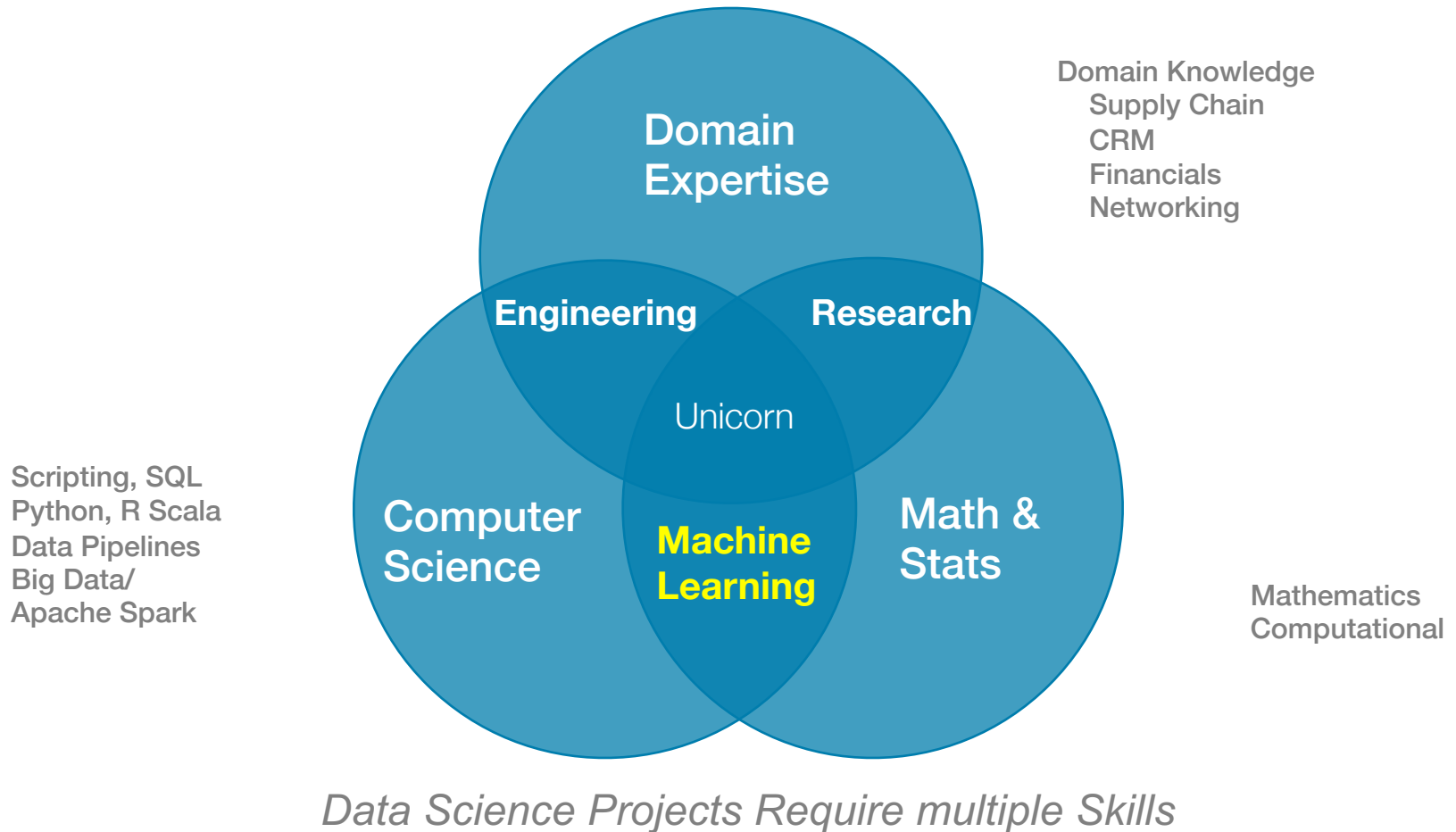
What is Machine Learning?

*“Computers that learn without being **explicitly programmed**”*

*“Using **algorithms** to understand patterns in data”*

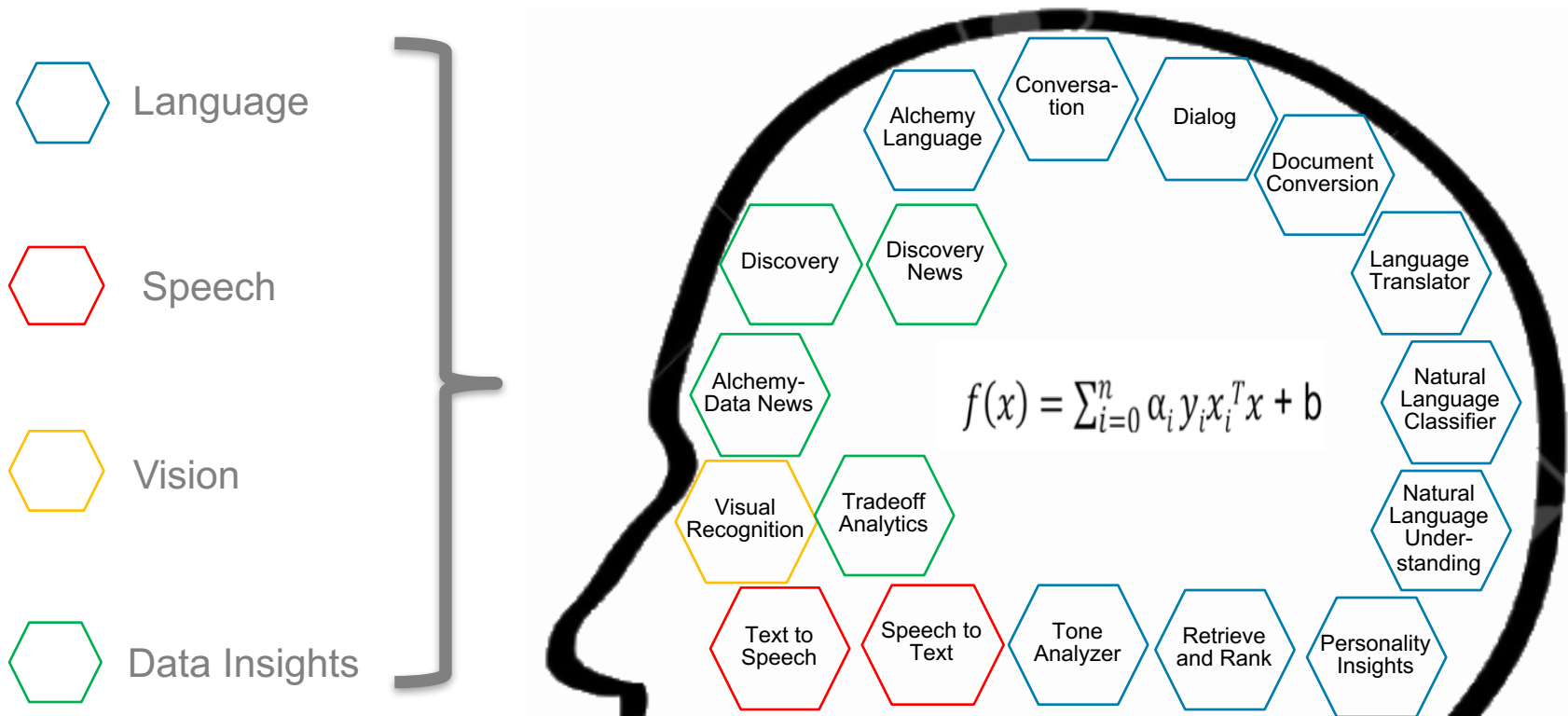


Machine Learning in Data Science Ecosystem....

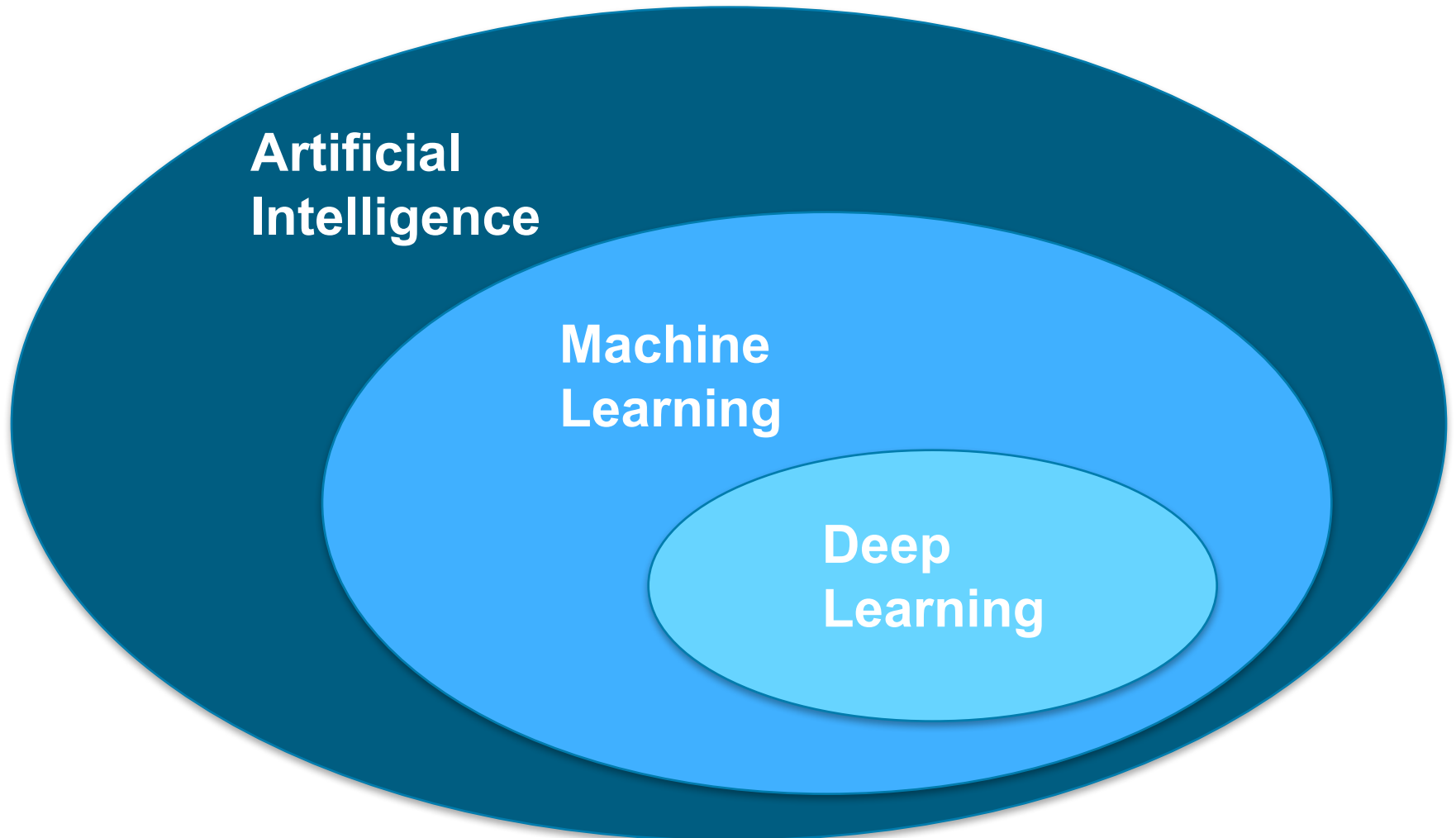


Machine Learning = Artificial Intelligence???

Data + Algorithms = Scored AI Models

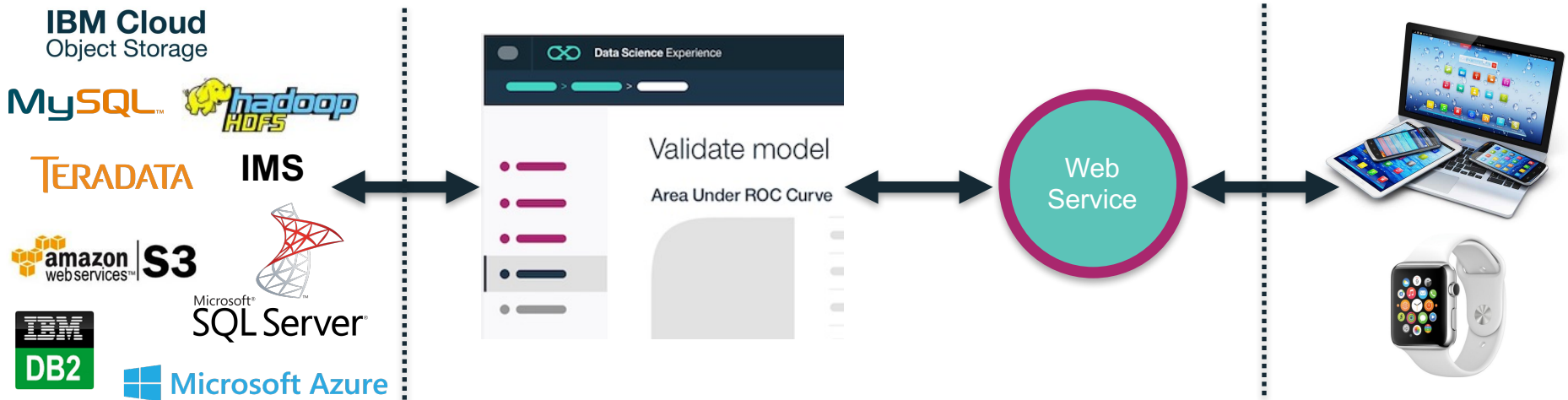


Understanding AI, ML & DL Relationship...





Introducing IBM Watson Machine Learning Accelerating Digital Deployments in Cloud...



Data Access:

- Easily connect to Behind-the-Firewall and Public Cloud Data
- Catalogued and Governed Controls through Watson Data Platform

Creating Models:

- Single UI and API for creating ML Models on various Runtimes
- Auto-Modelling and Hyperparameter Optimization

Web Service:

- Real-time, Streaming, and Batch Deployment
- Continuous Monitoring and Feedback Loop

Intelligent Apps:

- Integrate ML models with apps, websites, etc.
- Continuously Improve and Adapt with Self-Learning

IBM Watson Machine Learning in Data Science Experience



Data Scientist

APIs for Jupyter Notebooks

GA

The screenshot shows the IBM Data Science Experience Jupyter Notebook interface. The main area displays a tutorial titled "Deep Learning for Image Classification" with handwritten digits and their corresponding numerical labels: 7 → 7, 5 → 5, 8 → 8, 3 → 3, 2 → 2, and 4 → 4. Below the tutorial, there is a code cell with the following text:

```
Install the nolearn deep learning python library if you haven't
```

```
In [1]: #!pip install --user nolearn
```

```
In [ ]: import warnings
warnings.filterwarnings("ignore")
from sklearn.cross_validation import train_test_split
```

The right sidebar contains sections for "My Bookmarks", "Help", and "Community".

Model Builder GUI Lab 2

Open
Beta

The screenshot shows the IBM Data Science Experience Model Builder GUI. The top navigation bar includes "My Projects", "Demo project", and "Trained Model". The main area is titled "Select a technique" and features a sidebar with "Train" and "Evaluate" options. The "Train" section is active, showing a "Column value to predict (Label Col)" dropdown set to "PRODUCT_LINE" and a "Feature columns" dropdown set to "All (default)". Below these, there are three suggested techniques: "Binary Classification", "Multiclass Classification" (highlighted with a pink border), and "Regression". The "Multiclass Classification" box contains the text: "Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories." At the bottom, a "Validation Split" slider is shown with values for Train: 60, Test: 20, and Holdout: 20.

WML GUI - Train multiple models at same time Data Scientist

System tells you best model...

Open
Beta

IBM Data Science Experience

Projects
Catalog
Tools
Data Services
Community

US South
Docs

My Projects > Demo project > Naive

Select Data

Train

Evaluate

Select model

	ESTIMATOR TYPE	STATUS	PERFORMANCE	WEIGHTED TRUE POSITIVE RATE	WEIGHTED FALSE POSITIVE RATE	WEIGHTED PRECISION	WEIGHTED F MEASURE	WEIGHTED RECALL	LAST EVALUATION	ACTIONS
<input checked="" type="radio"/>	RandomForestClassifier	Trained & Evaluated	Fail	0.4329	0.12091	0.47587	0.37518	0.4329	10 Aug 2017, 2:43 PM	...
<input type="radio"/>	RandomForestClassifier	Trained & Evaluated	Fail	0.42051	0.1231	0.41262	0.36014	0.42051	10 Aug 2017, 2:42 PM	...
<input type="radio"/>	DecisionTreeClassifier	Trained & Evaluated	Fail	0.41019	0.11803	0.4318	0.37511	0.41019	10 Aug 2017, 2:36 PM	...
<input type="radio"/>	DecisionTreeClassifier	Trained & Evaluated	Fail	0.41019	0.11803	0.4318	0.37511	0.41019	10 Aug 2017, 2:38 PM	...
<input type="radio"/>	NaiveBayes	Trained & Evaluated	Fail	0.28128	0.15003	0.25212	0.25289	0.28128	10 Aug 2017, 2:37 PM	...
<input type="radio"/>	NaiveBayes	Trained & Evaluated	Fail	0.28128	0.15003	0.25212	0.25289	0.28128	10 Aug 2017, 2:41 PM	...

Close
Previous
Save

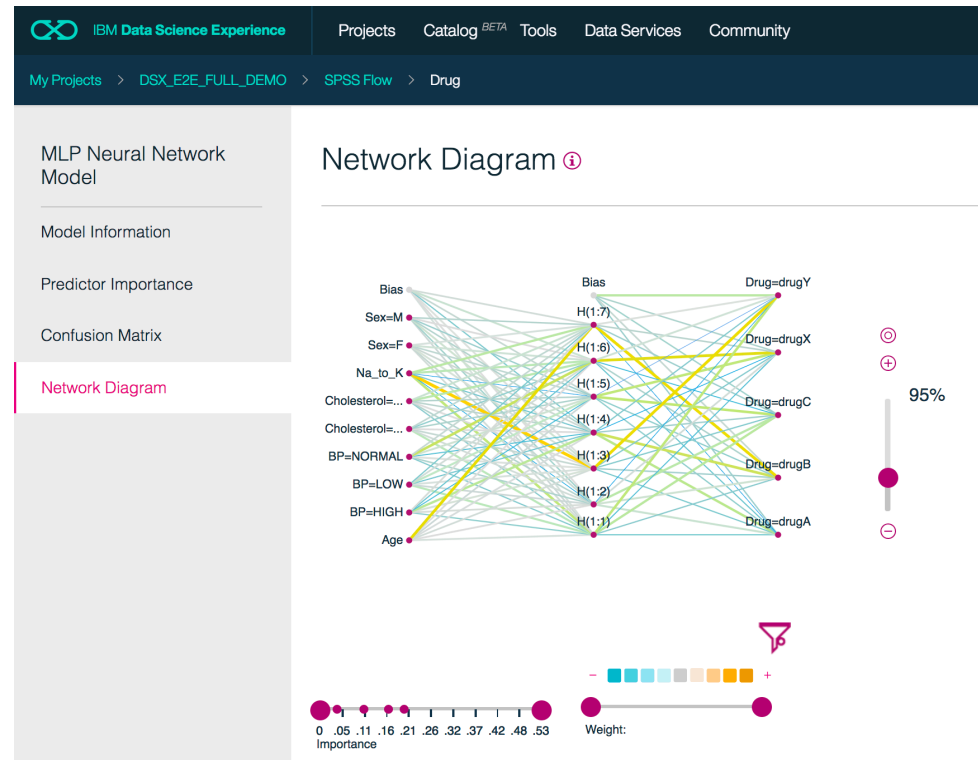
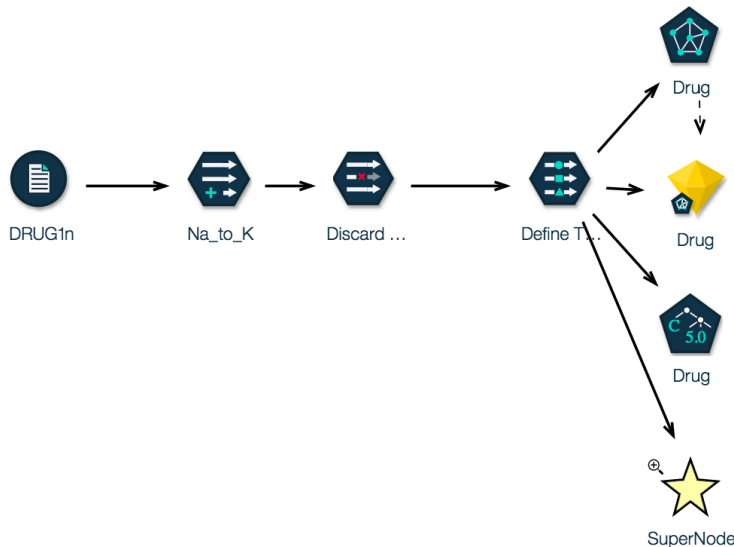
IBM Watson Machine Learning in Data Science Experience



Data Scientist

Flows

Open
Beta



Create Advanced Models without coding

Advanced Model Visualization: Easy to understand their performance

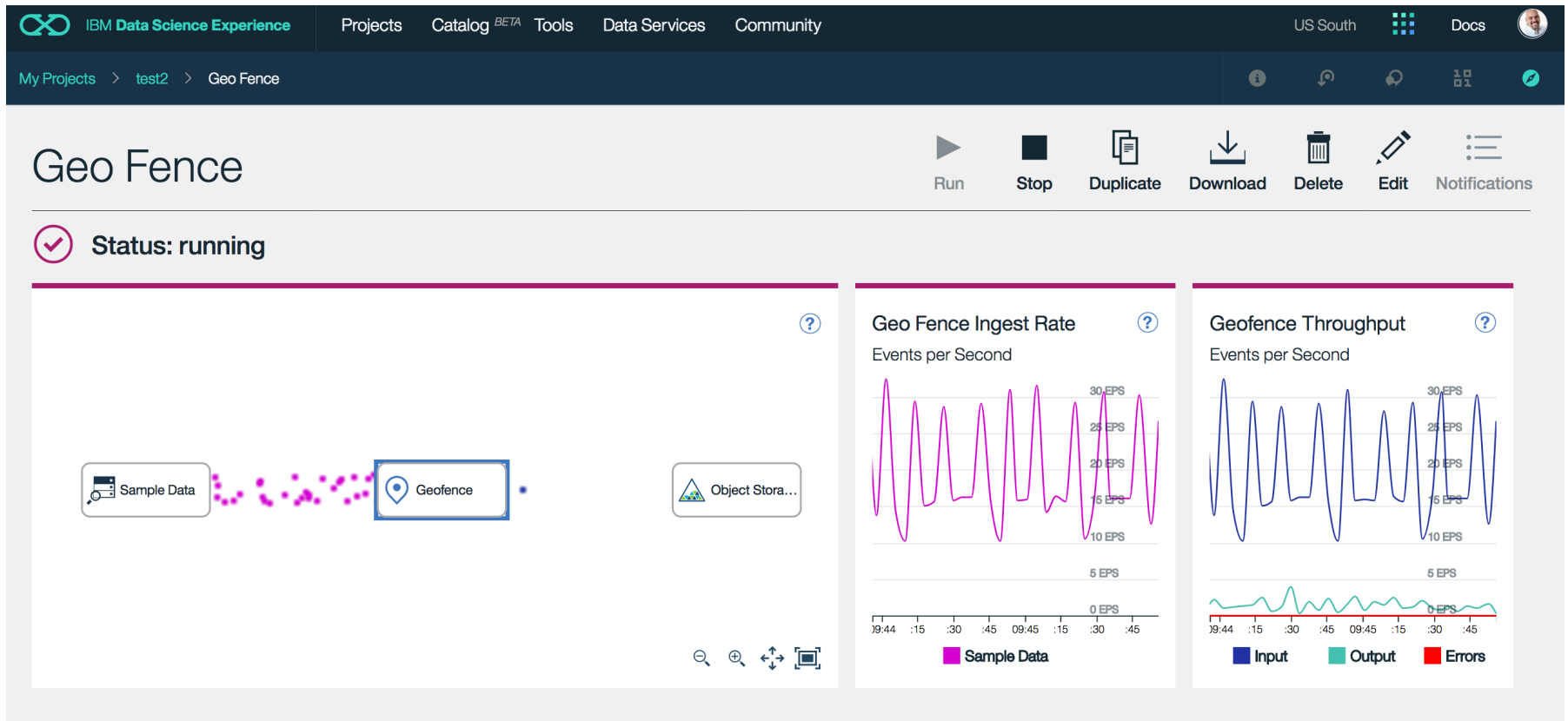
IBM Watson Machine Learning in Data Science Experience



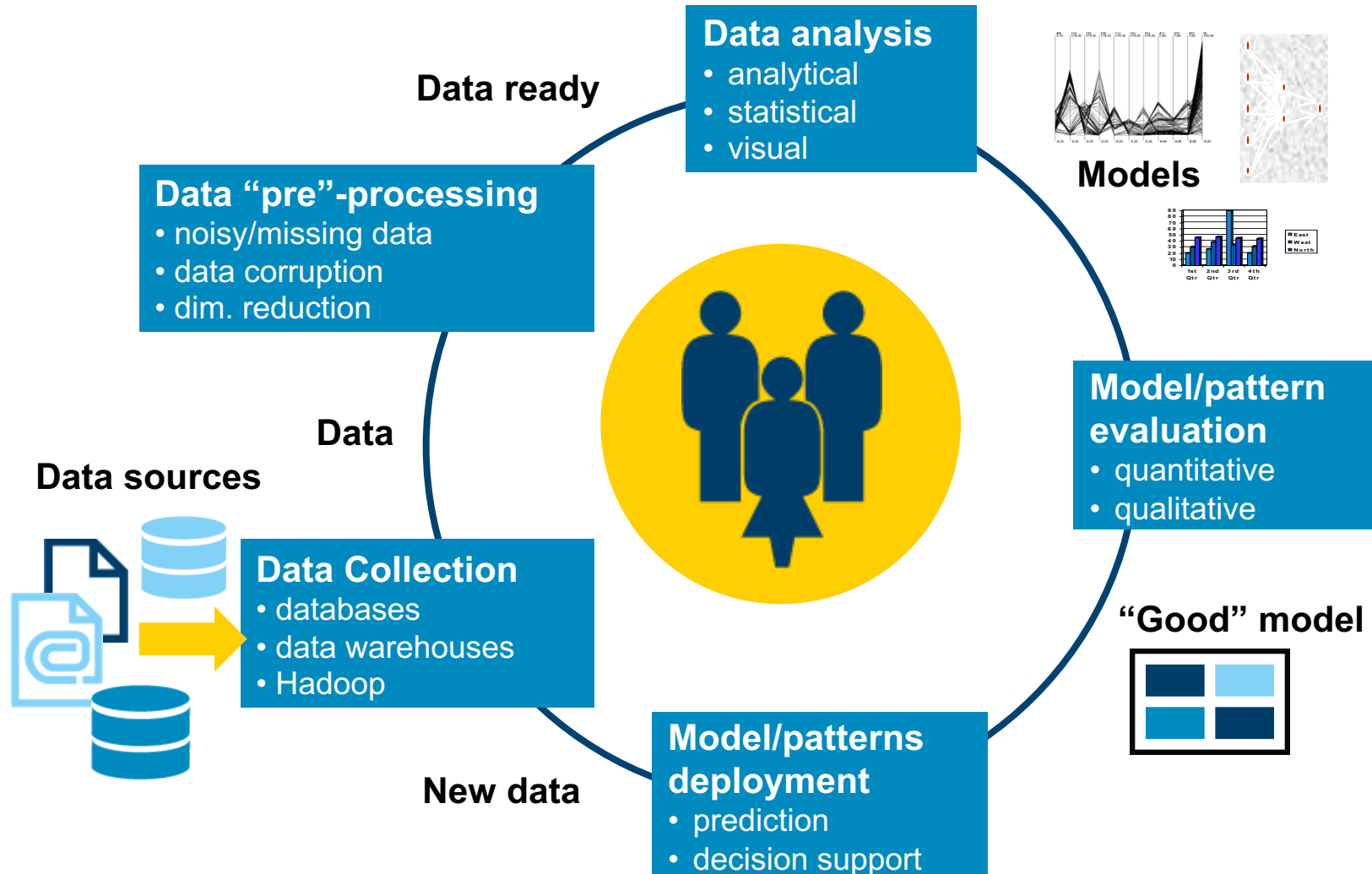
Data Scientist

Streaming Pipelines

Open
Beta



Data centric view of methodology



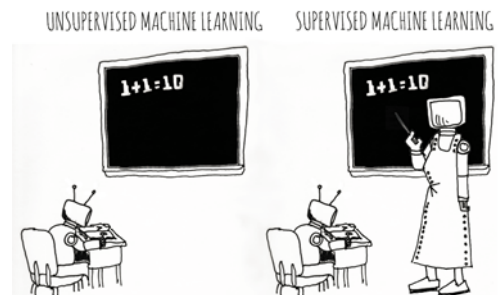
Categories of Machine Learning

■ Supervised learning

- The program is “trained” on a pre-defined set of “training examples”, which then facilitate its ability to reach an accurate conclusion when given new data
- The algorithm is presented with example inputs and their desired outputs (correct results)
- The goal is to learn a general rule that maps inputs to outputs

■ Unsupervised learning

- No labels are given to the learning algorithm, leaving it on its own to find structure (patterns and relationships) in its input
- Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning)

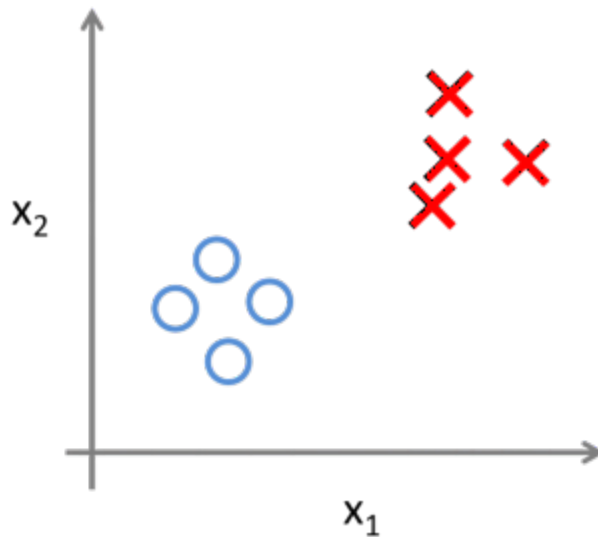


Categories of Machine Learning

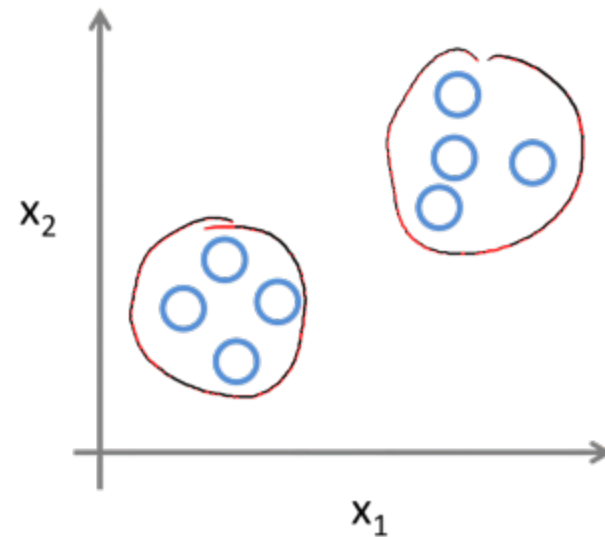
	Discrete Output	Continuous Output
Supervised Learning (require Ground-Truth)	<ul style="list-style-type: none"> • Classification (outcome is discrete) <ul style="list-style-type: none"> • Binary Classification <ul style="list-style-type: none"> • Linear Models (Logistic Regression) • Decision Trees • Naïve Bayes • Multi class Classification <ul style="list-style-type: none"> • Decision Trees • Naïve Bayes • K-NN 	<ul style="list-style-type: none"> • Regression <ul style="list-style-type: none"> - Linear - Ridge - Lasso • Decision Trees <ul style="list-style-type: none"> • Random Forest • Gradient Boosted Trees
Unsupervised Learning (no Ground-Truth data required)	<ul style="list-style-type: none"> • Clustering <ul style="list-style-type: none"> - k-means • FP-Growth 	<ul style="list-style-type: none"> • Clustering <ul style="list-style-type: none"> - k-means - Gaussian Mixture • Dimensionality Reduction <ul style="list-style-type: none"> - PCA - SVD

Supervised vs. Unsupervised Learning

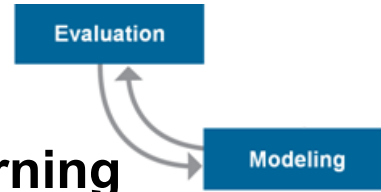
Supervised Learning



Unsupervised Learning



Training, testing, & validation sets



- During the model development process, supervised learning techniques employ **training** and **testing** sets and sometimes a **validation** set.
 - Historical data with known outcome (*target, class, response, or dependent variable*)
 - Source data randomly split or sampled... mutually exclusive records
- **Why?**
 - Training set → build the model (**iterative**)
 - Testing set → tune the parameters & variables during model building (**iterative**)
 - Assess model quality during training process
 - Avoid overfitting the model to the training set
 - Validation set → estimate accuracy or error rate of model (**once**)
 - Assess model's expected performance when applied to new data

Machine Learning – A more formal definition

Tom Mitchell of Carnegie Mellon University provides a widely quoted, more formal definition of machine learning

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E "



Q & A

