



Intro To Big Data Analytics and AI Flow

Intro To Big Data Analytics and AI

Impact of Data Analytics

Snow Leopards Demo

Big Data Analytics Hands-On-Lab

Data Analytics 101

The Evolution of Technologies

Analytics Architecture

Data Storage Considerations

Analytics Tools

AI 101

AI Hands-On-Lab

Deep Learning

Importance of Ethics

Impact of Data Analytics

Financial Services Use Cases

Effective customer engagement	Decision services management	Risk and revenue management	Risk and compliance management	Recommendation engine
Effective customer engagement Customer profiles Credit history Transactional data LTV Loyalty	Decision services management Customer segmentation CRM data Credit data Market data	Risk and revenue management Transaction data Demographics Purchasing history Trends	Risk and compliance management CRM Credit Risk Merchant records Products and services	Recommendation engine Clickstream data Products Services Customer service data
				
Customer analytics	Financial modeling	Risk, fraud, threat detection	Credit analytics	Marketing analytics
Customer 360 degree evaluation Customer segmentation Reduced customer churn Underwriting, servicing and delinquency handling Insights for new products	Commercial/retail banking, securities, trading and investment models Decision science, simulations and forecasting Investment recommendations	Real-time anomaly detection Card monitoring and fraud detection Security threat identification Risk aggregation	Enterprise DataHub Regulatory and compliance analysis Credit risk management Automated credit analytics	Recommendation engine Predictive analytics and targeted advertising Fast marketing and multi-channel engagement Customer sentiment analysis
Faster innovation for a better customer experience	Improved consumer outcomes and increased revenue	Enhanced customer experience with machine learning	Transform growth with predictive analytics	Improved customer engagement with machine learning

Health & Life Sciences Use Cases

DNA sequences	Real world analytics	Image deep learning	Sensor data	Social data listening
FAST-Q BAM SAM VCF Expression	HL7/CCD 837 Pharmacy Registry EMR	MRI X-RAY CT Ultrasound	Readings Time series Event data	Social media Adverse events Unstructured
				
Genomics and precision medicine Single cell sequencing Biomarker, genetic, variant and population analytics ADAM and HAIL on Databricks	Clinical and claims data Claims data warehouse Readmission predictions Efficacy and comparative analytics Prescription adherence Market access analysis	GPU image processing Graphic intensive workloads Deep learning using Tensor Flow Pattern recognition	IoT device analytics Aggregation of streaming events Predictive maintenance Anomaly detection	Social analytics Real-time patient feedback via topic modelling Analytics across publication data
Faster innovation for drug development	Improved outcomes and increased revenue	Diagnostics leveraging machine learning	Predictive analytics transforms quality of care	Improved patient communications and feedback

Media & Entertainment Use Cases

Personalized recommendations	Effective customer retention	Information optimization	Inventory allocation	Consumer engagement analysis
Customer profiles Viewing history Online activity Content sources Channels	Customer profiles Online activity Content distribution Services data	Consumption logs Clickstream and devices Marketing campaign responses	Transactions Subscriptions Demographics Credit data	Content metadata Ratings Comments Social media activity
 Content personalization Personalized viewing and engagement experience Click-path optimization Next best content analysis Improved real time ad targeting	 Customer churn prevention Quality of service and operational efficiency Market basket analysis Customer behavior analysis Click-through analysis	 Recommendation engine Ad effectiveness Content monetization Fraud detection Information-as-a-service High value user engagement	 Predictive analytics Predict audience interests Network performance and optimization Pricing predictions Nielsen ratings and projections Mobile spatial analytics	 Sentiment analysis Demand-elasticity Social network analysis Promotion events time-series analysis Multi-channel marketing attribution
Faster innovation for customer experience	Improved consumer outcomes and increased revenue	Enhance user experience with machine learning	Predictive analytics transforms growth	Improved consumer engagement with machine learning

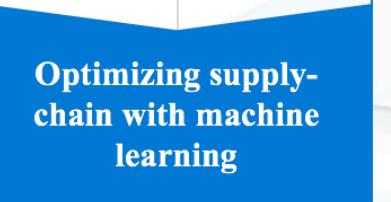
Retail Use Cases

Recommendation engine	Effective customer engagement	Inventory optimization	Inventory allocation	Consumer engagement
<p>Customer profiles Shopping history Online activity Social network analysis</p> 	<p>Shopping history Online activity Floor plans App data</p> 	<p>Demand plans Forecasts Sales history Trends Local events/weather patterns</p> 	<p>Demographics Buyer perception Consumer research Market/competitive analysis</p> 	<p>Historical sales data Price scheduling Segment level price changes</p> 
<p>Next best and personalized offers</p> <p>Customer 360/consumer personalization Right product, promotion, at right time Multi-channel promotion</p>	<p>Store design and ergonomics</p> <p>Path to purchase In-store experience Workforce and manpower optimization</p>	<p>Data-driven stock, inventory, ordering</p> <p>Predict inventory positions and distribution Fraud detection Market basket analysis</p>	<p>Assortment optimization</p> <p>Economic modelling Optimization for foot traffic, Online interactions Flat and declining categories</p>	<p>Real-time pricing optimization</p> <p>Demand-elasticity Personal pricing schemes Promotion events Multi-channel engagement</p>
<p>Faster innovation for customer experience</p>	<p>Improved consumer outcomes and increased revenue</p>	<p>Omni-channel shopping experience with machine learning</p>	<p>Predictive analytics transforms growth</p>	<p>Improved consumer engagement with machine learning</p>

Advertising & Marketing Tech Use Cases

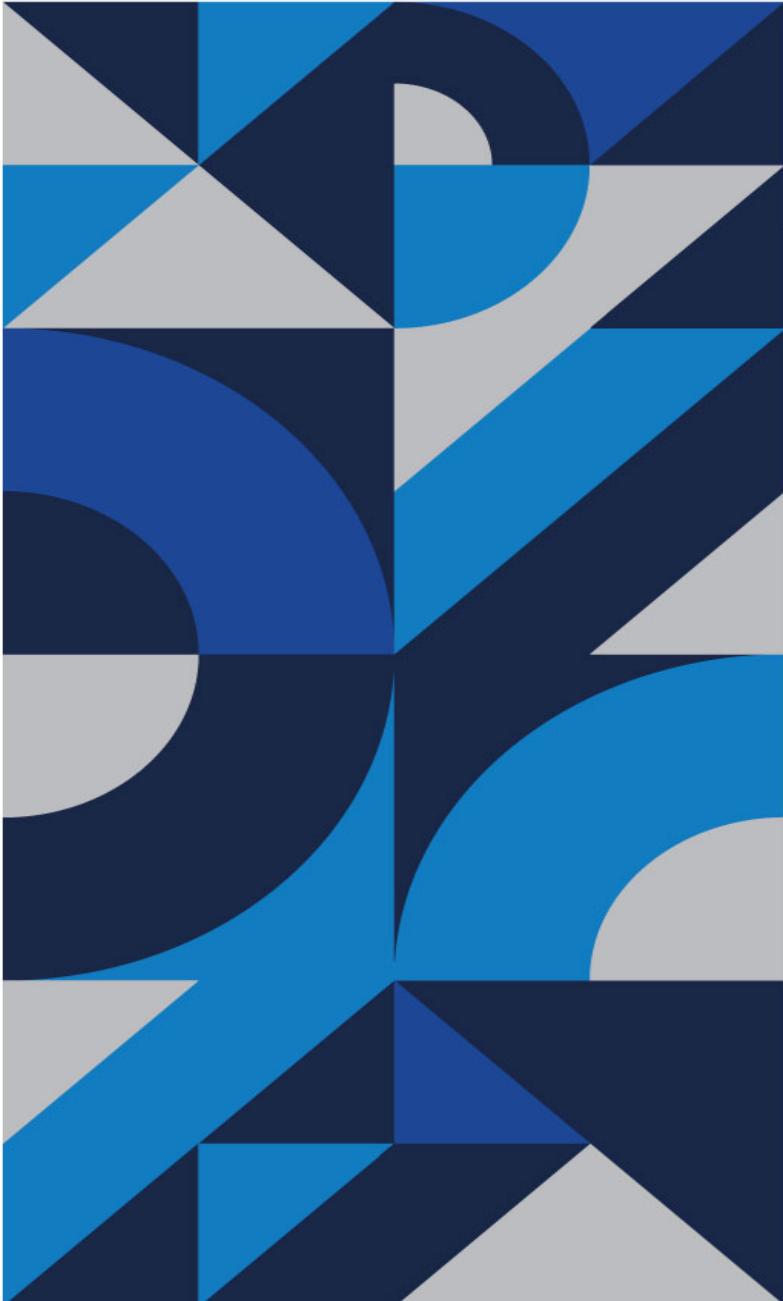
Effective customer engagement	Recommendation engine	Risk and fraud analysis	Campaign reporting analytics	Brand promotion and customer experience
Customer profiles Online history Transaction data Loyalty	Customer segmentation CRM data Credit data Market data	Transaction data Demographics Purchasing history Trends	CRM Merchant records Products Services Marketing data	Social media Online history Customer service data
				
Customer value analytics Customer 360, segmentation aggregation and attribution Audience modelling/index report Reduce customer churn Insights for new products Historical bid opportunity as a service	Next best and personalized offers Right product, promotion, at right time Real time ad bidding platform Personalized ad targeting Ad performance reporting	Risk and fraud management Real-time anomaly detection Fraud prevention Customer spend and risk analysis Data relationship maps	Sales and campaign optimization Optimizing return on ad spend and ad placement Multi-channel promotion Ideal customer traits Optimized ad placement	Sentiment analysis Opinion mining/social media analysis Deeper customer insights Customer loyalty programs Shopping cart analysis
Faster innovation for customer growth	Improved outcomes and increased revenue	Risk management with machine learning	Predictive analytics transforms growth	Improved customer engagement with machine learning

Oil & Gas Use Cases

Upstream optimization, maximize well life Field data Asset data Demographics Production data 	Grid operations, asset inventory optimization Sensor stream data UAVs images Inventory data Production data 	Supply-chain optimization Transaction data Demographics Purchasing history Trends 	Risk optimization Sensor stream data Transport Retail data Grid production data Refinery tuning parameters 	Recommendations engine Clickstream data Products Services Market data Competitive data Demographics 
Digital oil field/oil production Production optimization Integrate exploration and seismic data Minimize lease operating expenses Decline curve analysis 	Industrial IoT Pipeline monitoring Preventive maintenance Smart grids and microgrids Grid operations, Field Service Asset performance as a Service 	Supply-chain optimization Trade monitoring, optimization Retail mobile applications Vendor management - construction, transportation, truck & delivery optimization 	Safety and security Real-time anomaly detection Predictive analytics Industrial safety Environment health and safety 	Sales and marketing analytics Fast marketing and multi-channel engagement Develop new products and monitor acceptance of rates Predictive energy trading Deep customer insights 
Faster innovation for revenue growth	Improved outcomes and increased revenue	Optimizing supply-chain with machine learning	Predictive analytics transforms safety and security	Improved customer engagement with machine learning

Security Use Cases

Security controls to leverage all data	Actionable threat intelligence	Risk and fraud analysis	Compliance management	Identity and access management for analytics
Firewall/network logs Apps Data access layers	Firewall/network logs Network flows Authentications	Firewall/network logs Web/app logs Social media content	Firewall/network logs Web Applications Devices OS	Files Tables Clusters Reports Dashboards Notebooks
				
Intrusion detection and predictive analytics Prevention of DDoS attacks Threat classifications Data loss/anomaly detection in streaming Cybermetrics and changing use patterns	Security intelligence Real-time data correlation Anomaly detection Security context, enrichment Offence scoring, prioritization Security orchestration	Fraud detection and prevention e-Tailing Inventory monitoring Social media monitoring Phishing scams Piracy protection	Security compliance reporting Ad-hoc/historic incident reports SOC/NOC dashboards Deep OS auditing Data loss detection in IoT User behavior analytics	Fine-grained data analytics security Role-based access controls Auditing and governance File integrity monitoring Row level and column level access permissions
Prevent complex threats with machine learning	Faster innovation for threat prevention	Risk management with machine learning	Transform security with improved visibility	Limit malicious insiders to transform growth



Snow Leopards Demo



NEWS

[Home](#) | [Video](#) | [World](#) | [US & Canada](#) | [UK](#) | [Business](#) | [Tech](#) | [Science](#) | [Stories](#) | [Entertainment](#)[Asia](#) | [China](#) | [India](#)

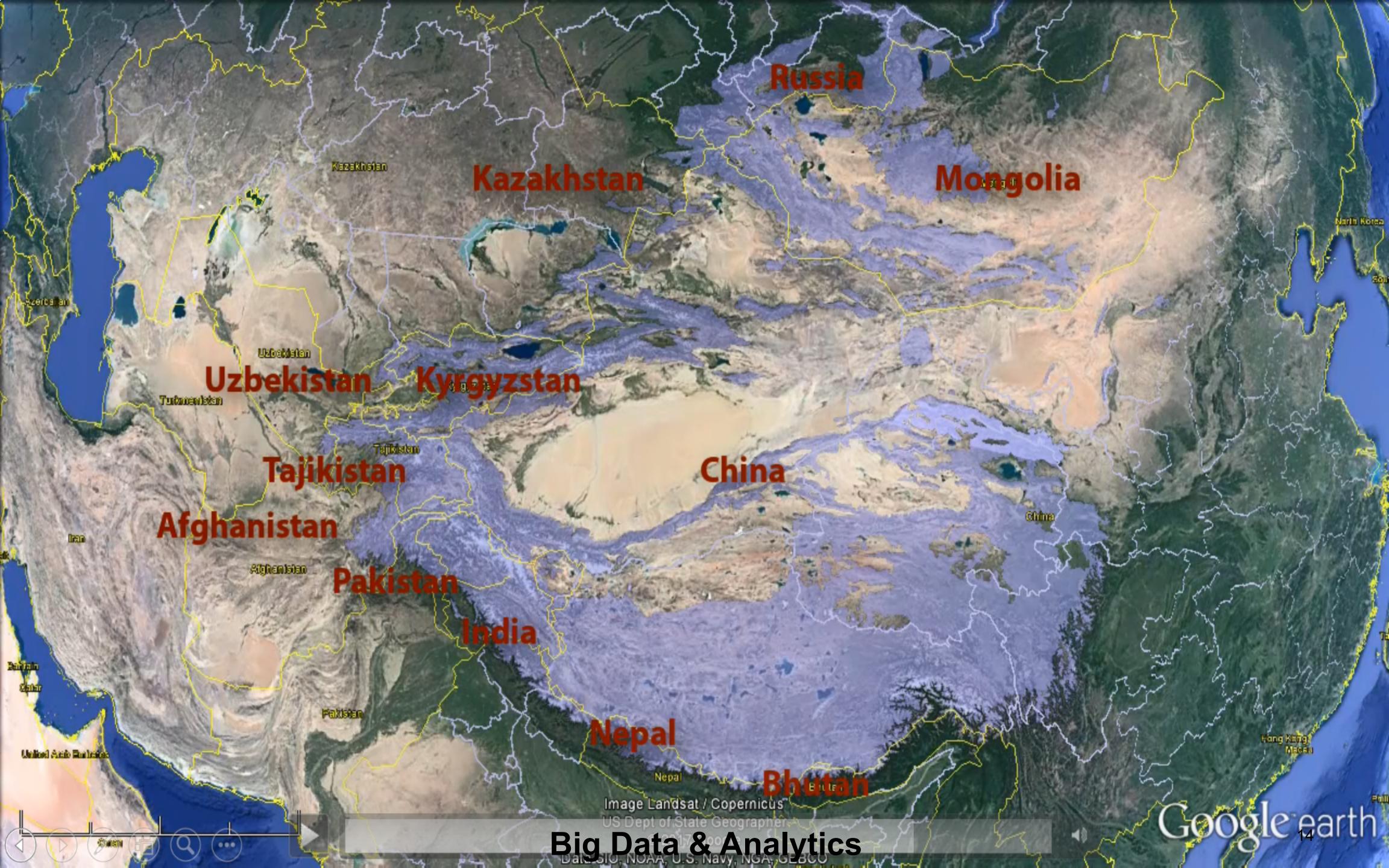
Snow leopard no longer 'endangered'

① 14 September 2017



Statement on IUCN Red List Status Change of the Snow Leopard

The Snow Leopard Trust, one of the leading conservation organizations working to protect this cat, opposes the IUCN's decision to change the snow leopard's Red List status from 'Endangered' to 'Vulnerable'.



Big Data & Analytics

Image Landsat / Copernicus
US Dept of State Geographer
Date: 01/01/2017
Data: USGS, NOAA, U.S. Navy, NGA, GEDCO

Google earth
14



Snow
Leopard
Trust



Classifying all 1.3 million images will take 20k man hours







High level library for distributed machine learning

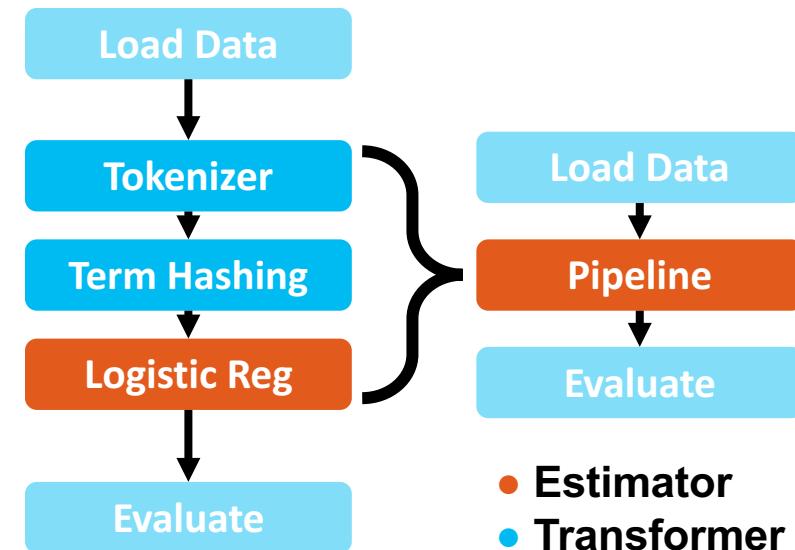
More general than SciKit-Learn

All models have a uniform interface

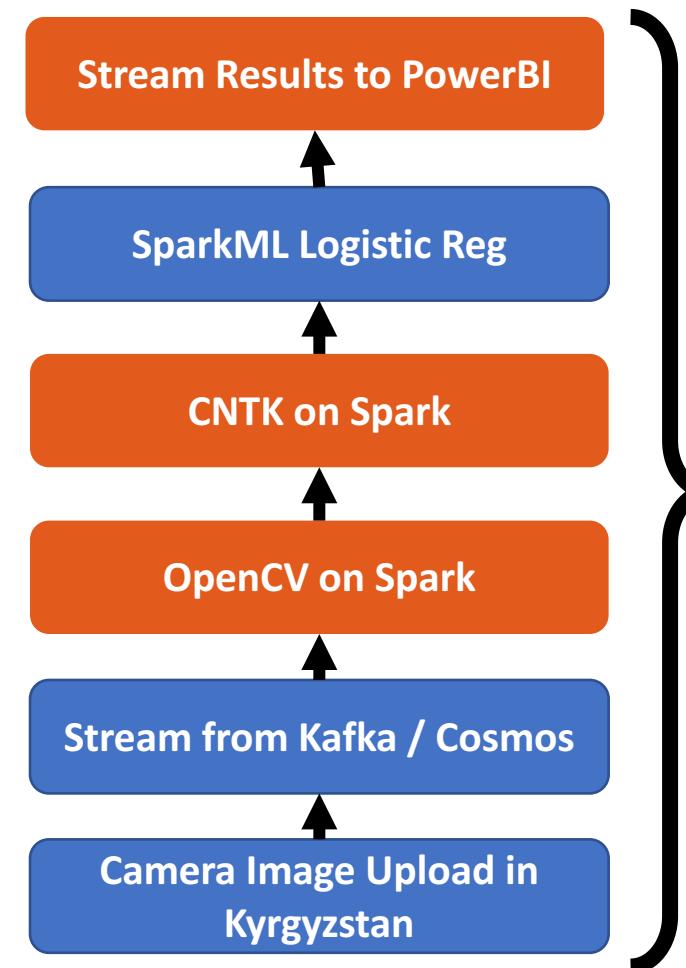
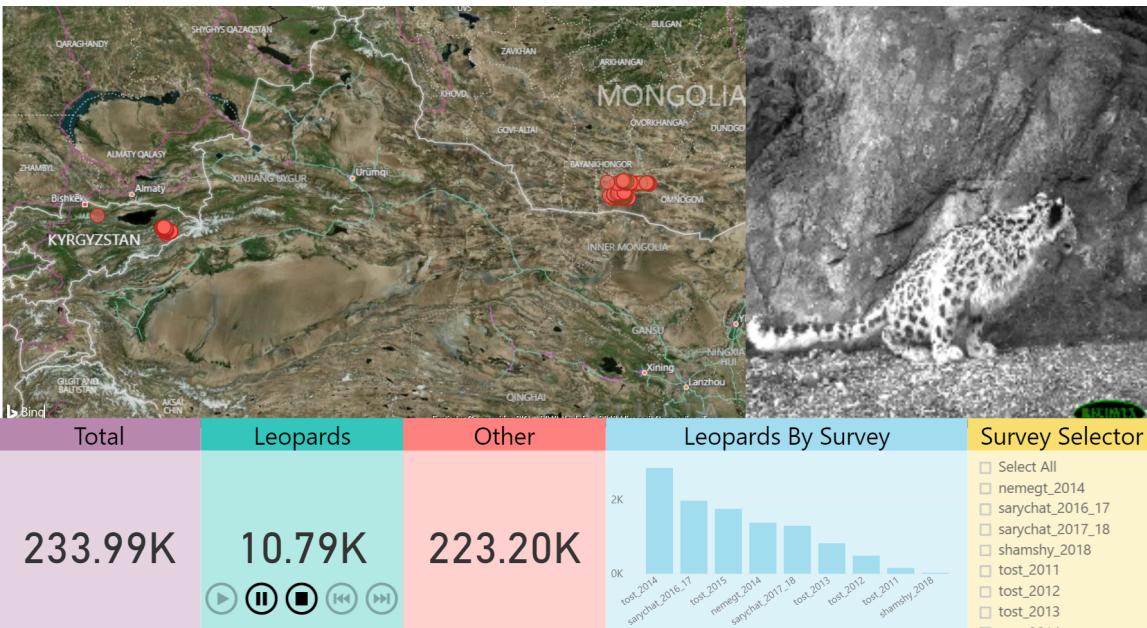
Can compose models into complex pipelines

Can save, load, and transport models

```
data = spark.read.csv("hdfs://...")  
train, test = data.randomSplit([.5,.5])  
model = LogisticRegression().fit(train)  
predictions = model.transform(test)
```



Real Time Monitoring with Dashboards





Basics of Data Analytics

4 Types of Data Analytics

Descriptive

Insight based on past info

What happened?

Example:
Used in standard report generation and aggregated totals.

Diagnostic

Examines the cause of past results

Why did it happen?

Example:
Used in interactive dashboards and variance analyses to examine the causes.

Predictive

Provides foresight for the future by identifying patterns in historical data.

What will happen?

Example:
Used to predict likelihood of a customer renewing a subscription.

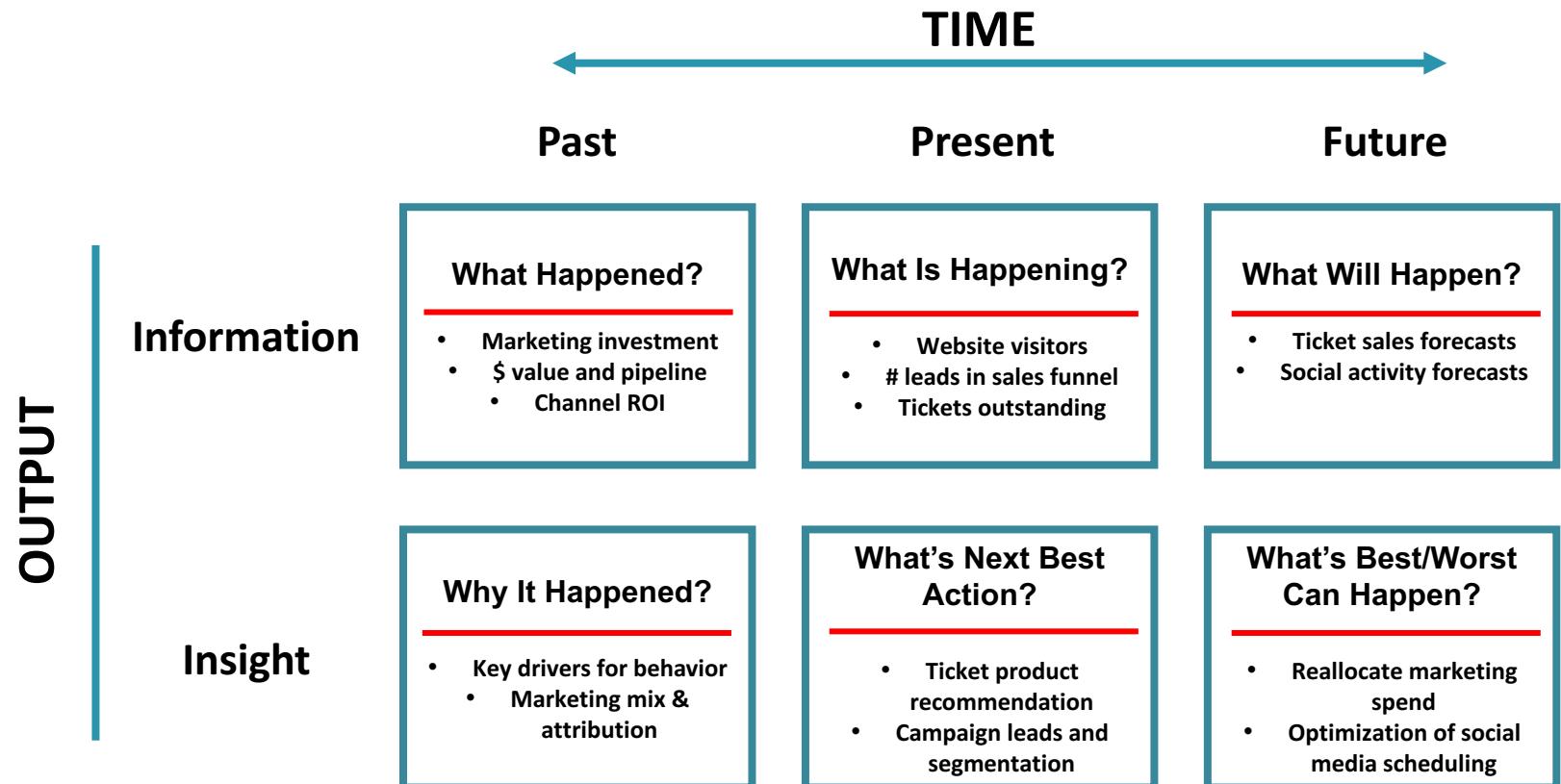
Prescriptive

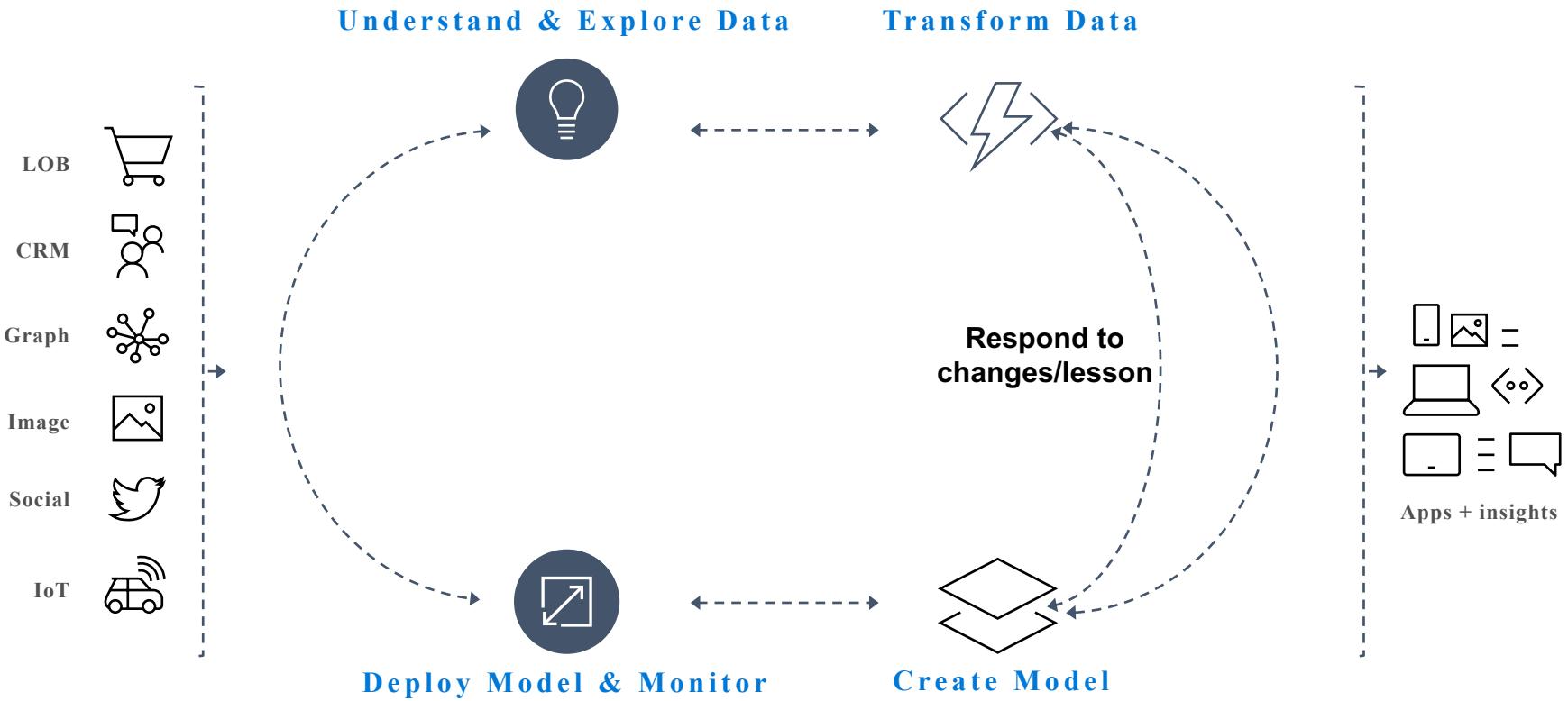
Assists in identifying best option for the desired and optimal outcome.

What should we do?

Example:
Used to identify actions to reduce the collection period for accounts receivables.

Key Analytics Objectives





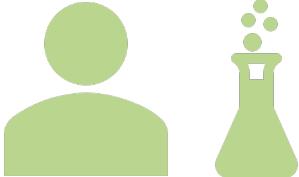


Big Data Analytics Roles



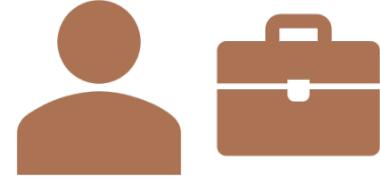
Data Engineer

- ✓ Transform data
- ✓ Create and schedule batch
- ✓ Streaming ETL jobs
- ✓ Manage platforms
- ✓ Agile development
- ✓ Knowledge of:
 - ✓ Platforms
 - ✓ Java, C++
 - ✓ Scripting
 - ✓ Python, ...



Data Scientist

- ✓ Explore data
- ✓ Create and deploy analytical models
- ✓ Transform data
- ✓ Perform advanced analytics tasks
- ✓ Knowledge of:
 - ✓ Statistics
 - ✓ Machine Learning
 - ✓ Linear Algebra
 - ✓ R, ...



Business Analyst

- ✓ Write SQL queries
- ✓ Generate reports
- ✓ Explore data
- ✓ Analyze and visualize data
- ✓ Knowledge of basic analysis tools such as:
 - ✓ Analytics IDEs
 - ✓ Power BI
 - ✓ Tableau



Supervised vs Unsupervised

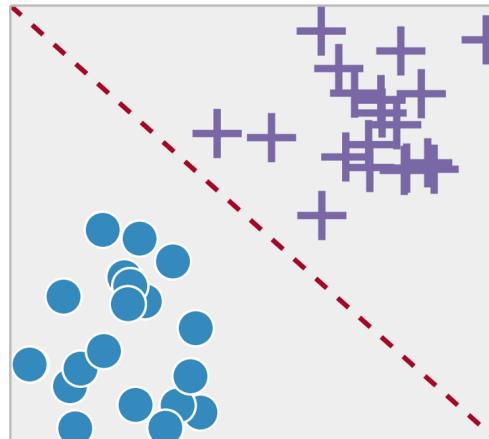
Supervised vs Unsupervised

Supervised

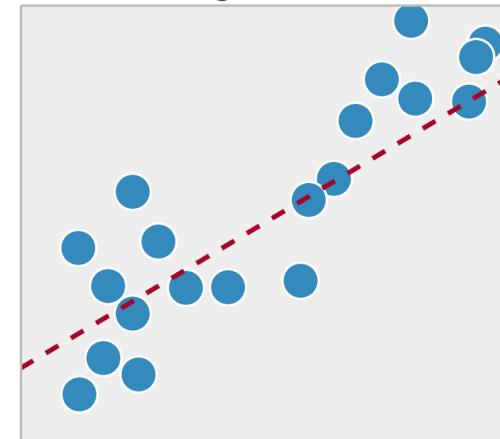
Leveraging the prior knowledge of the true outcomes as a target and basis for learning.

Example:
Classification or Regression Models

Classification



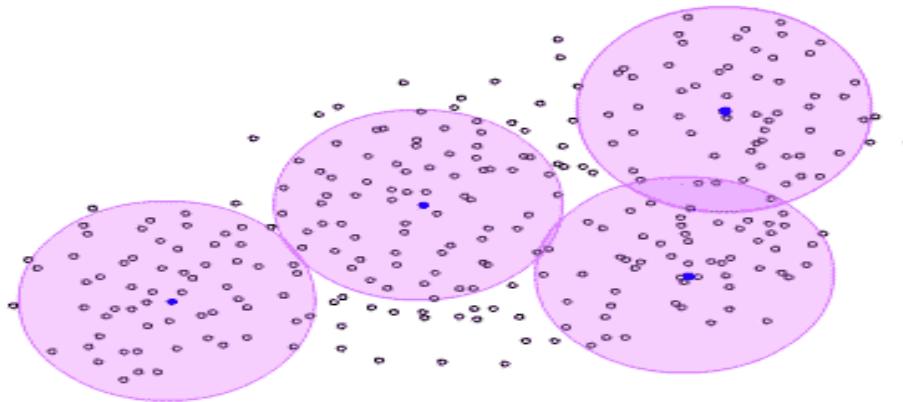
Regression

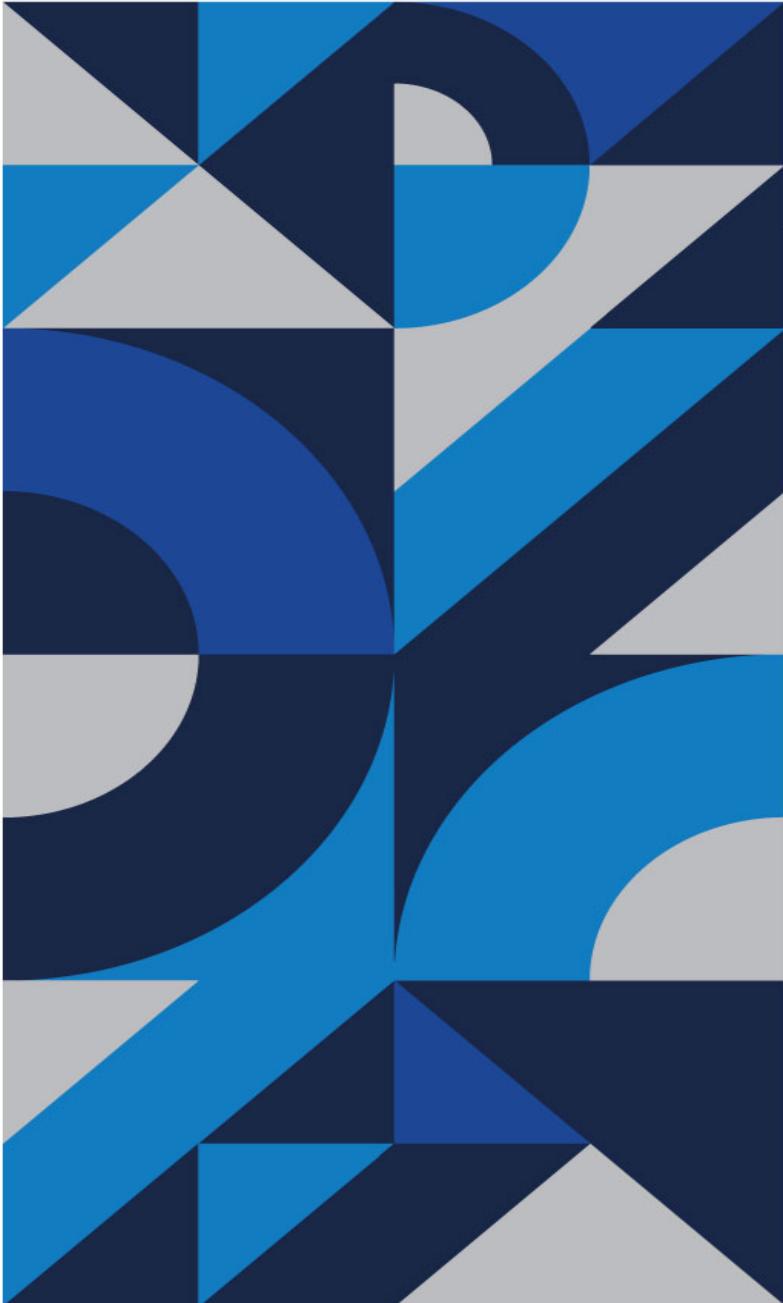


Unsupervised

Learning the inherent structure of the data without using explicitly-provided labels.

Example:
Clustering





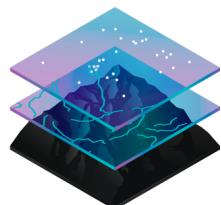
Key Analytics Genres

Key Analytics Genres

Time Series/Forecasting



Spatial Analysis



Log Analytics



Bioinformatics



Consumer Insights



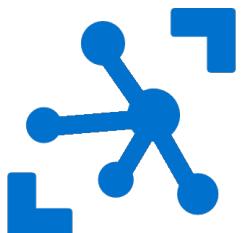
Video/Image



Search | Text | NLP



IoT/Streaming



Audio



The Evolution of Technologies

Each business is a victim of Digital Darwinism, the evolution of consumer behavior when society and technology evolve faster than the ability to exploit it. Digital Darwinism does not discriminate. Every business is threatened.

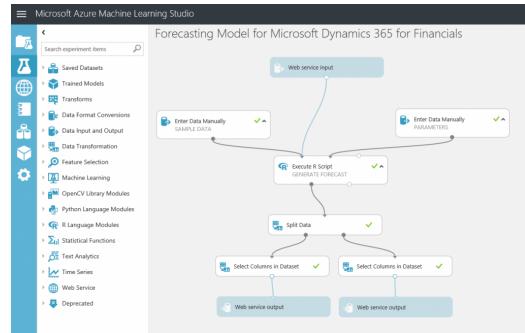
– Brian Solis

Democratizing Analytics

Can we develop new tools and applications to allow common business personnel to leverage advanced analytics.

```
data = spark.read.csv("hdfs://...")  
train, test = data.randomSplit([.5,.5])  
model = LogisticRegression().fit(train)  
predictions = model.transform(test)
```

GUIs



New tools leverage simple interfaces with user-friendly features to increase adoption of tools and services.

Programmatic

For full flexibility and control, programmatic approach to applications is the primary method for advanced analytics development.

Example of Analytics Services Offering

DaaS

Data as a service is an offering that gathers, processes and provides relevant data on demand.

Example:
Infogroup
Axiom

SaaS

Software as a service provides a web based distribution of software where the provider is responsible for hosting management.

Example:
Snowflake
Databricks

AaaS

Analytics as a Service provides prebuilt analytical models and content that can be “called” or leveraged on demand.

Example:
Azure Cognitive Services
AWS ML APIs
Google ML APIs

IaaS

Insights as a service aim to provide direct insights and visualizations reflecting results from analyses from customer submitted data.

Example:
Fanalytical

Key Technologies and Applications

Machine Learning

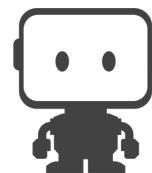


Deep Learning



TensorFlow





DataRobot



Gartner Report





Cloud vs On-premise

Cloud

Data is Gold

Infrastructure is currency

Cloud providers are the new age banks.

Give away software, just provide the data.

On-premise

Responsible for infrastructure and software to perform analytics.

Data storage Considerations

Machine Learning Data

Most machine learning analytical tasks require data in a relational structure.

Most analytical technologies have their custom concept of a relational table or “dataframe” to perform operations.



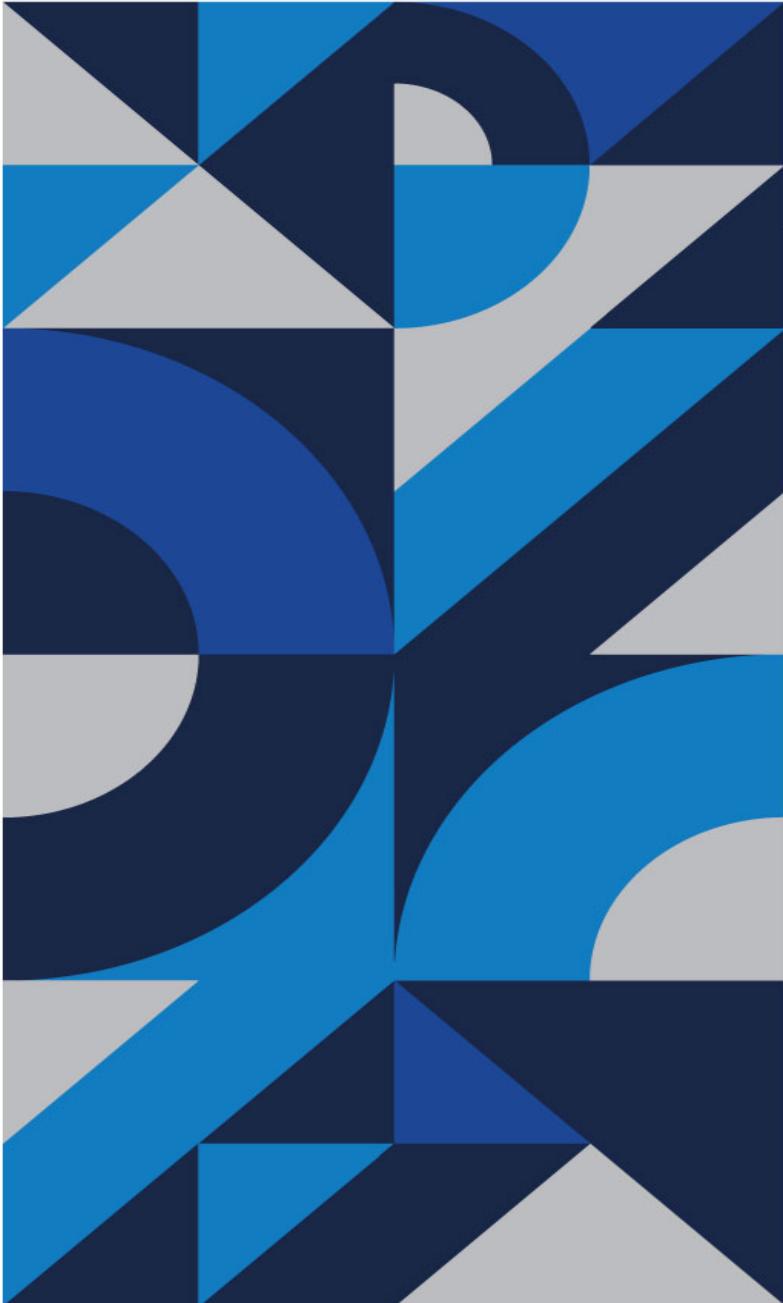
Deep Learning Data

Most deep learning analytical tasks iterate over a list of objects/files referenced in a folder or key-value store.

Many applications make this process seamless, by only requiring the location/folder of the corresponding files.

Audio, video, text, and image files are very common for DL tasks.



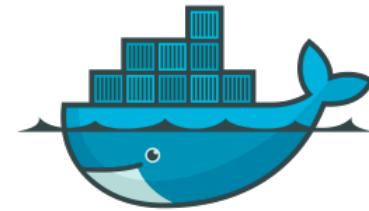


Analytics Tools

IDEs



Code Repo & DevOps



docker



GitHub



GitLab

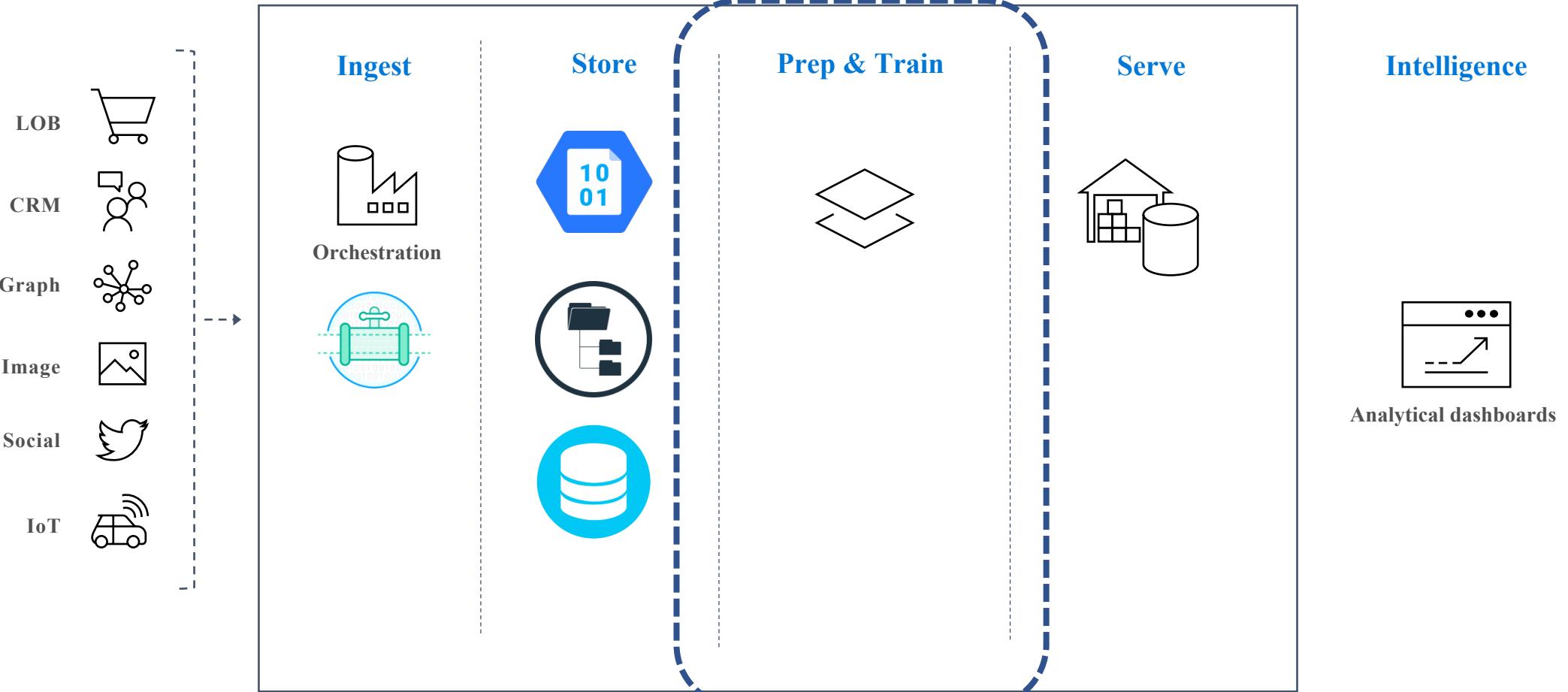


kubernetes



Databricks, SAS Viya, Cloud ML & AI

Architecture Diagram Icons



Characteristics

Distributed In-memory Compute

SAS R&D technology

Memory persistence

Notes

Very efficient for psuedo-large datasets.
Memory intensive.
~100x > MapReduce.
Core facilities: SAS Procs, ML, Streaming
New SAS CASL language and SAS BASE compatible.

Icons

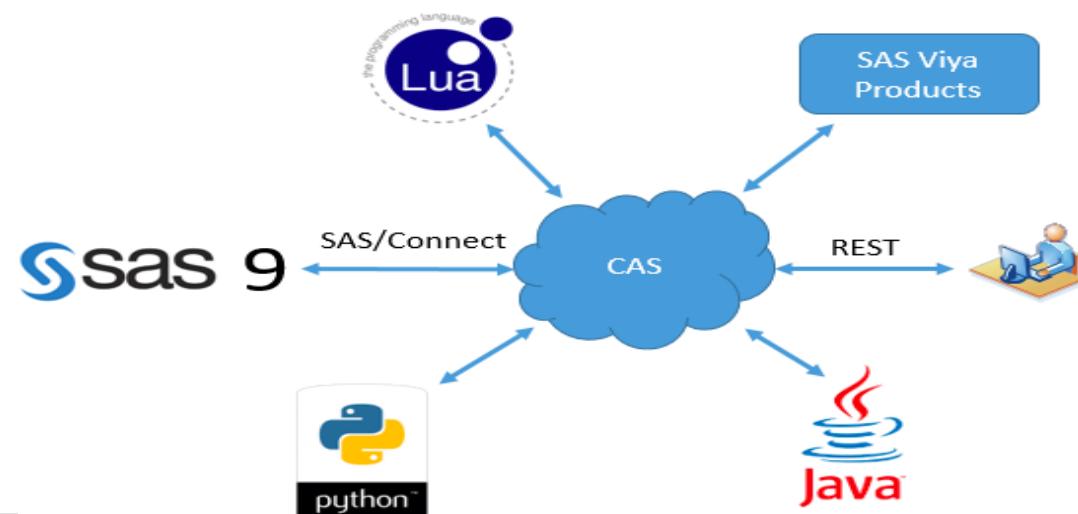


SAS® Viya®



\$\$\$\$

Cost



Big Data & Analytics

Characteristics

SaaS-like offering of Apache Spark

Spark inventors are founders

Collaborative tool

\$\$

Cost

Notes

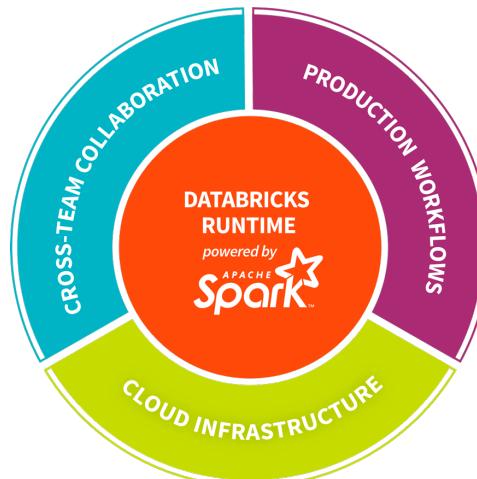
Data engineers, data scientists and data analysts.
Model pipelining.
100% Apache Spark compliant.
Spark streaming.
Powerful Notebook style editing.
Databricks File System.
Built-in visualizations

Icons



databricks™

Azure Databricks



Characteristics

A suite of custom, optimized libraries for ML & AI.

Designed to simplify code effort and increase adoption.

Flexible deployment and dev infrastructure.



Cost

Notes

Each cloud provider has own offerings for analytics and AI.

Azure has a DL framework CNTK.

Google has DL framework Tensorflow.

AWS and Azure have proprietary streaming services.

AI models via APIs.

Requires learning curve.

Icons



Amazon
Machine
Learning





Introduction to Artificial Intelligence

AI

Any technique that enables a computer to mimic human behavior.

Note:

In analytics space, AI and Deep Learning are the same. True AI would incorporate application development with analytics incorporated.

Self-driving Cars

Vehicle Cybersecurity

Driving Assistant

Vehicle systems

Customer Service

Call Analytics

Call Intent Discovery

Service Chat Bots

Analytics

Image Recognition

Conversational Analytics

Survey Analytics (NLP)



Hands-On-Lab: Azure Cognitive Services



What is Deep Learning?

Machine Learning

Algorithms that can learn from data without relying on rules-based programming.



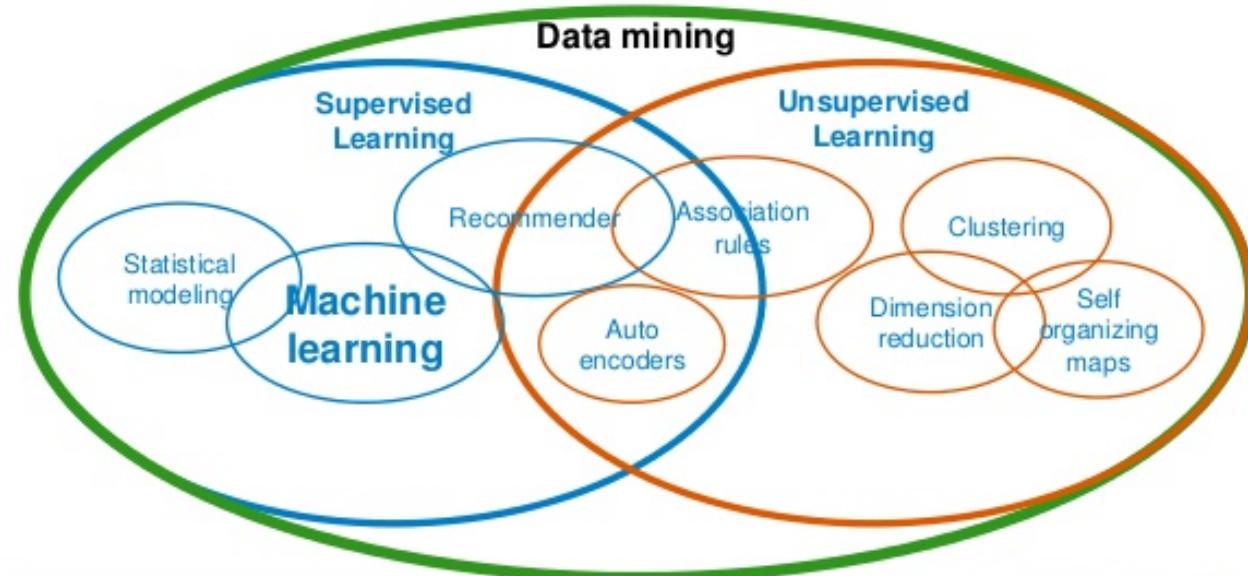
Statistical modeling

Formalization of relationships between variables in the data in the form of mathematical equations.

“Glorified statistics”

... is what Robert Tibshirani, a statistician and machine learning expert at Stanford, calls machine learning.

MACHINE LEARNING AND SOME OTHER TERMS YOU OFTEN HEAR



- ❑ Machine learning treats an algorithm like a black box, as long it works.
- ❑ It is generally applied to high dimensional data sets, the more data you have, the more accurate your prediction is.

Accuracy > Theory

Machine Learning Components

Supervised Learning

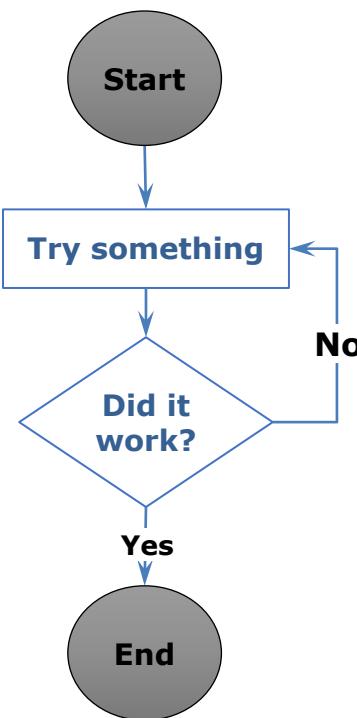
Learning with a labeled training set

Unsupervised learning

Discovering patterns in unlabeled data

Reinforcement learning

Learning based on feedback or reward



Linear Regression

SVM

PCA

Random **Forest**

Decision *Trees*

Artificial Neural Networks

K-means

Artificial Neural Network (ANN)



- ANN is about:
 - One input
 - One output and
 - Multiple fully-connected hidden layers in-between.
- The more layers the network has, the higher-level features it will learn.
- Most deep networks use, for hidden layer, the activation function

$$g(x) =: \text{ReLU} = \max(0, x)$$

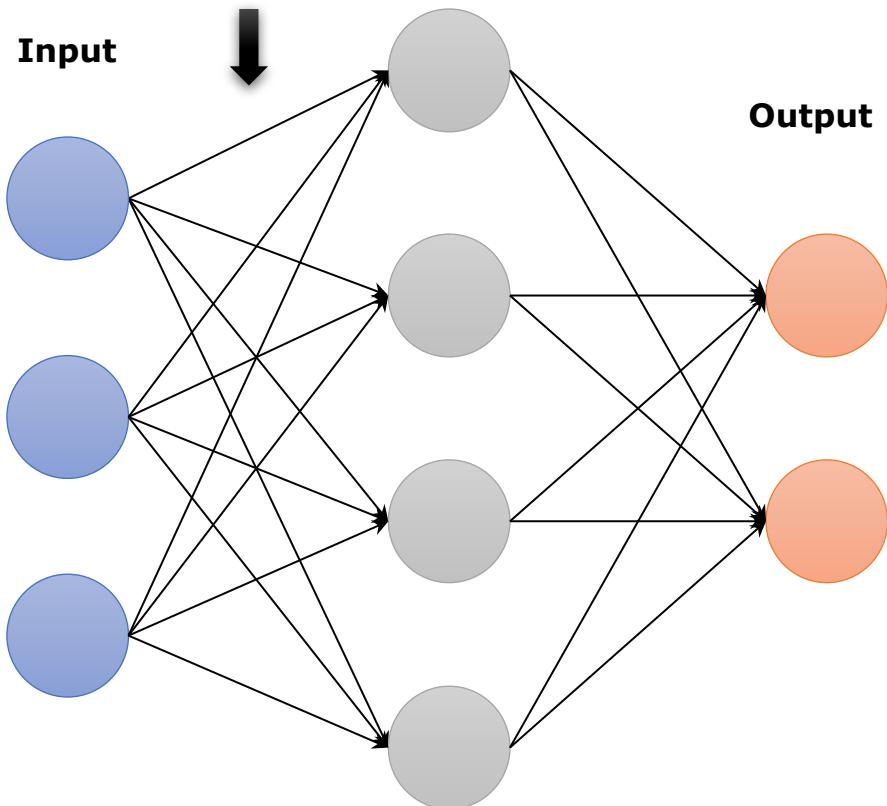
Activation Function:

$$g(x)$$

Hidden

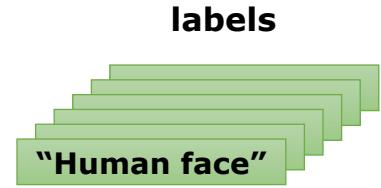
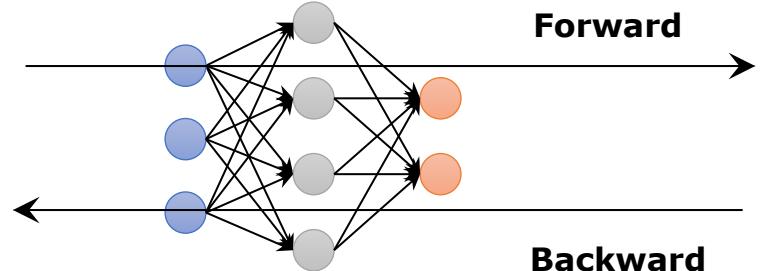
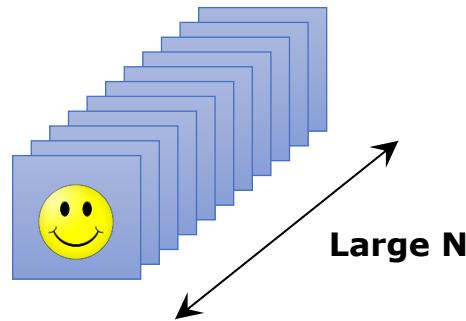
Input

Output

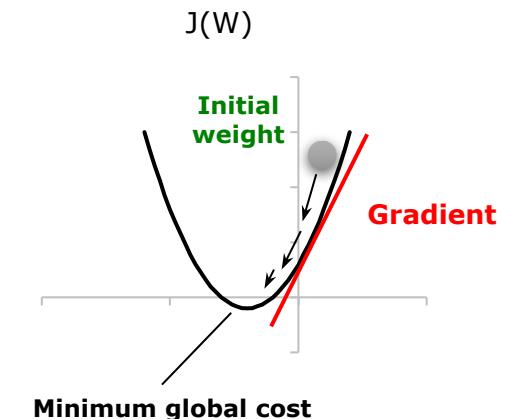
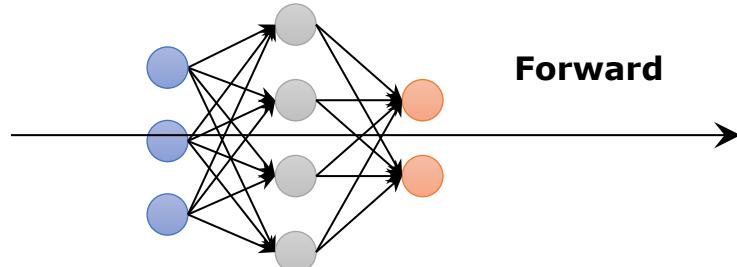
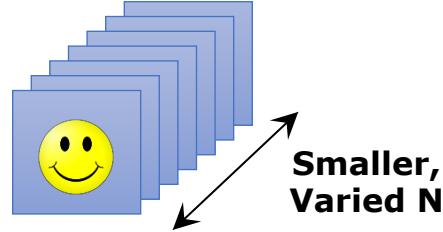


Deep Neural Network

Training

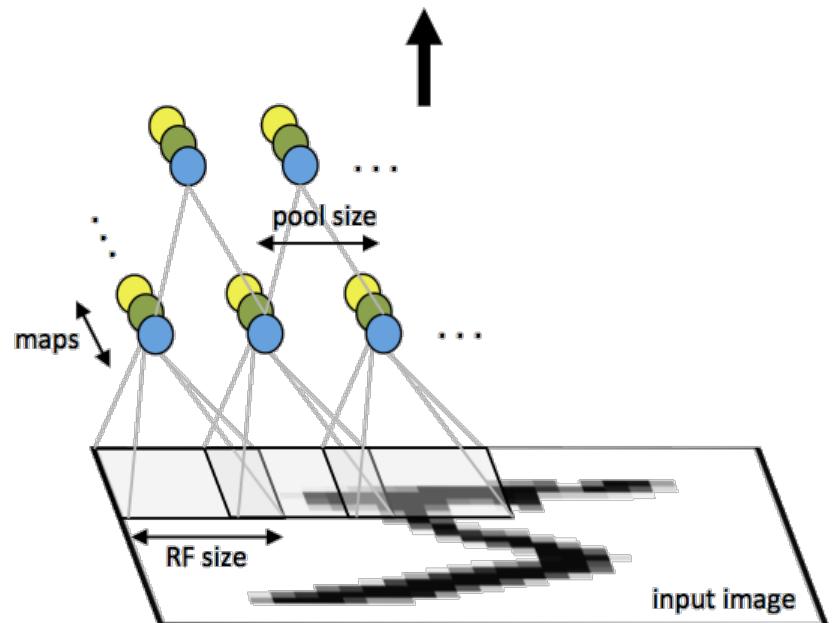


Inference

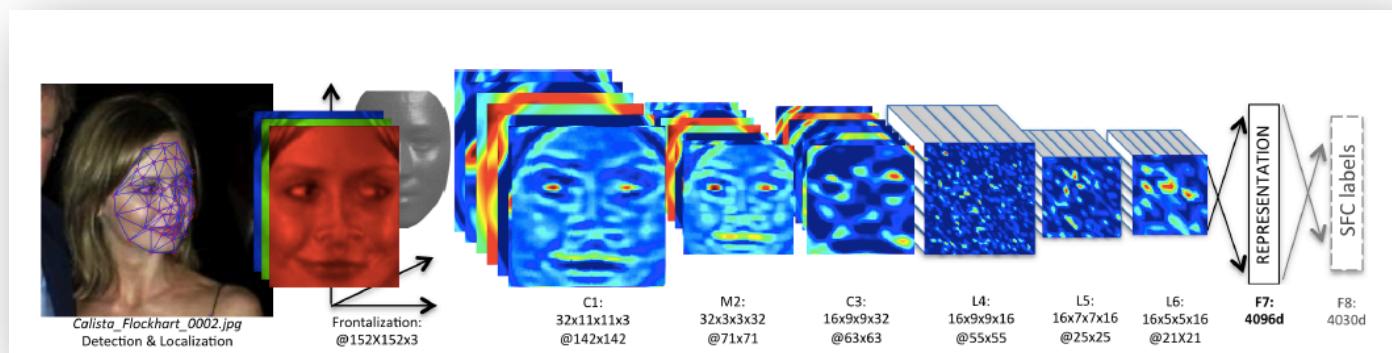


DNN learn by generating an error signal that measures the difference between the predictions of the network and the desired values and then using the error signal to change the weights (or parameters) so that predictions get more accurate.

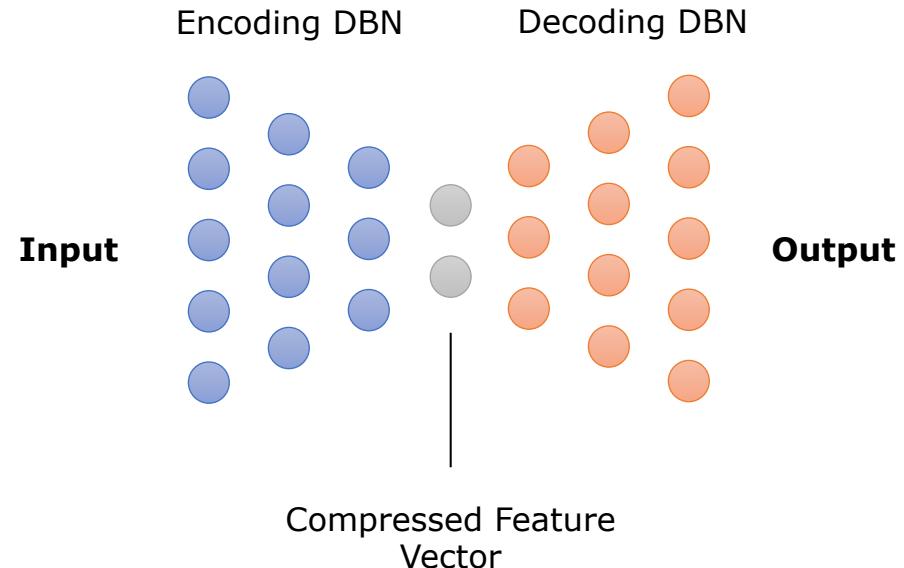
Convolutional Neural Network (CNN)



- ❑ A CNN consists of a number of convolutional and subsampling layers optionally followed by fully connected layers.
- ❑ The input to a convolutional layer is a $m \times m \times r$ image where m is the height and the width of the image and r is the number of channels, e.g. an RGB image has $r = 3$.
- ❑ Every layer takes a 3D volume of numbers and outputs a 3D volume of numbers.



Deep Autoencoder



- ❑ A deep autoencoder is composed of two, symmetrical deep-belief networks that typically have four or five shallow layers representing the encoding half of the net, and second set of four or five layers that make up the decoding half.

- ❑ The layers are restricted Boltzmann machines, the building blocks of deep-belief networks.
- ❑ Use Cases:
 - Image Search
 - Data Compression
 - Topic discovery and information retrieval (IR)

Recurrent Neural Network (RNN)

- ❑ RNNs are general computers which can learn algorithms to map input sequences to output sequences.
- ❑ The output vector's contents are influenced by the entire history of inputs.
- ❑ Adaptive Robotics, handwriting recognition, image classification, speech recognition, stock market prediction

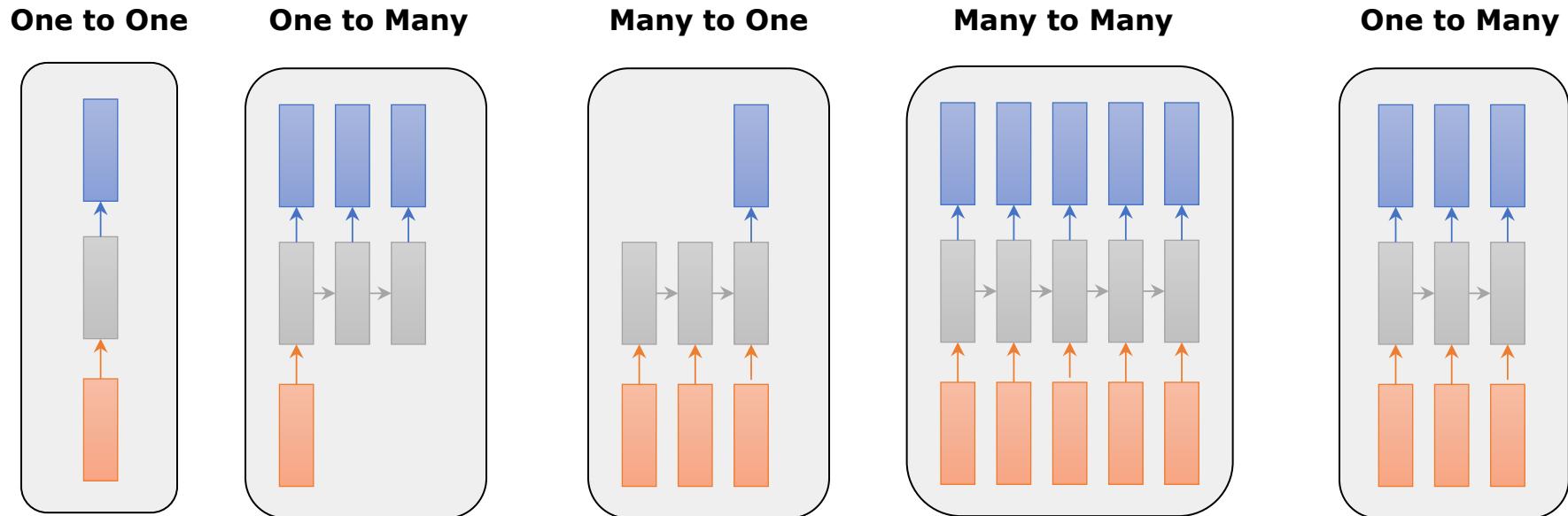


Image Captioning

Some of the captions can be unbelievably good ...



Two pizzas sitting on top of a
stove top oven



A group of young people playing a game of Frisbee



... and some less!

What computer sees

Image classification:
82% cat
15% dog
2% hat
1% mug

Image **captioning** = Image **recognition** + Language **generation**

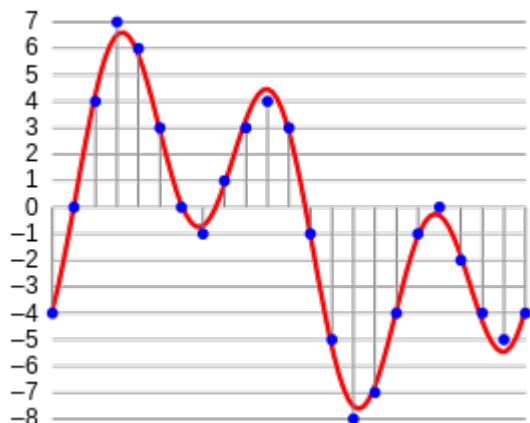
Notes

.wav, mp3, wma

Time x Amplitude

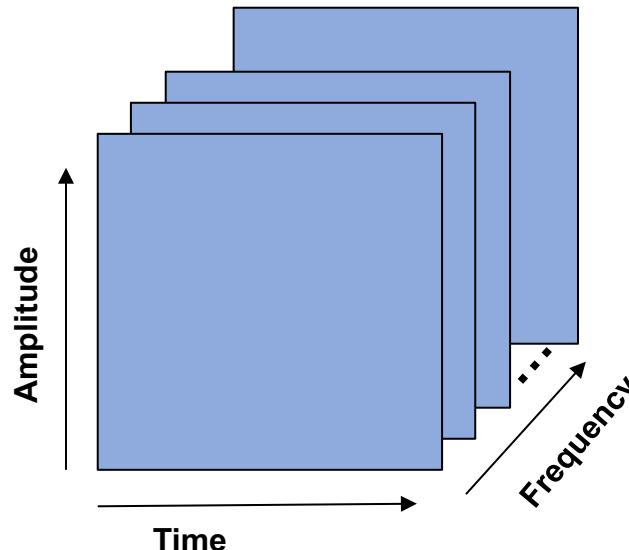
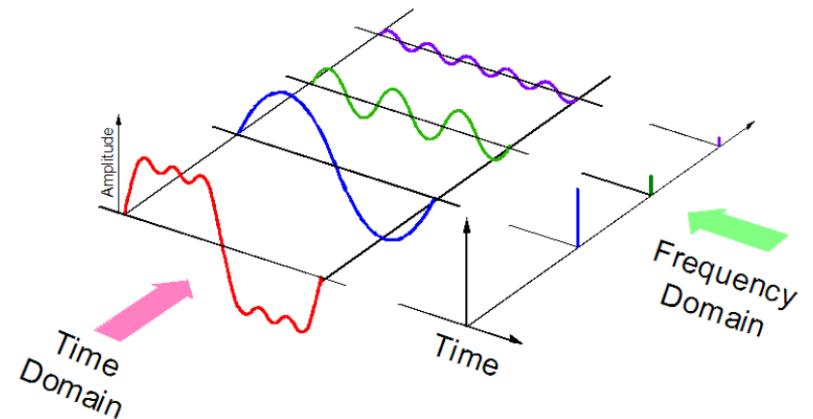
Transform file into a tensor
(multi-dim array)

Sampling



&

Re-representation



Audio Tensor

Big Data & Analytics

Notes

**Images have 3 dimensions:
height, width, color**

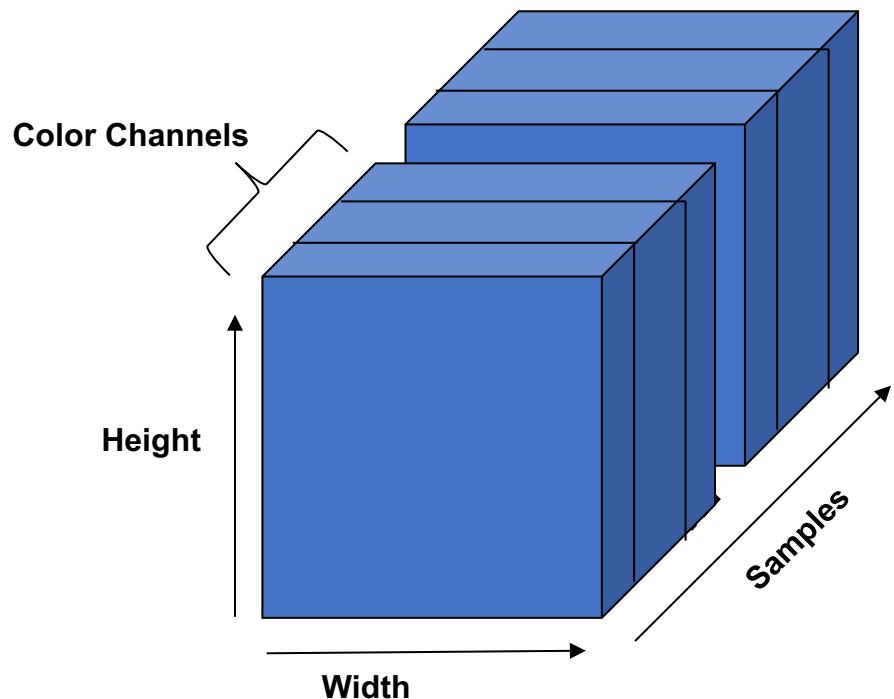
Time x Amplitude

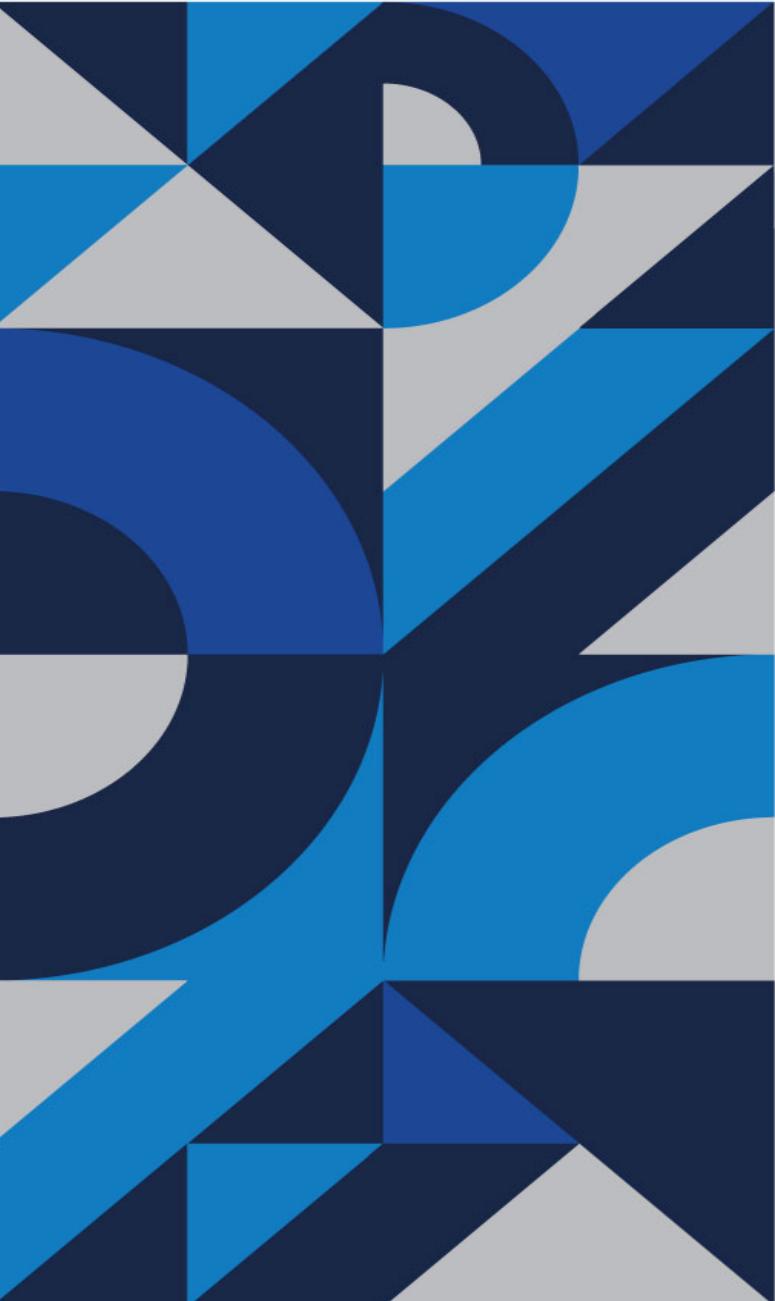
**Transform file into a tensor
(multi-dim array)**

Example Photo



Image Tensor





Deep Learning Frameworks

Tensorflow/Keras

Developed by Google
Most popular on Github
Well documented and followed on StackOverflow.
Tough to debug
Static language

CNTK

Developed by Microsoft
Supports multiple MSFT internal services
Less popular in community.
Deprecated

PyTorch

Developed by Facebook
Gaining momentum for adoption.
MSFT announced internal adoption over CNTK
Known for simple debugging feature and simple syntax.
Dynamic language

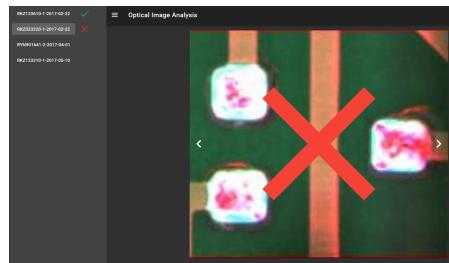


Image Classification

Image Classification Examples

Image Classification

Jabil, Printed Circuit Board Defect Detection



Schneider Electric, Circuit Breaker Defect Detection



Object detection

Shell, Counting Grey Stock of Pipes for Inventory Management



Darlie, Retail Store Inspection



Image segmentation

Land O'Lakes, Sustainability Farming, Map Labeling



Mask RCNN



Dice coefficient
waterways: 0.42
fieldborders: 0.86



Unet

Dice coefficient
waterways: 0.6
fieldborders: 0.11

Image Similarity

Brillio - Which images are similar to the query image?





Importance of Ethics

The New York Times

Facial Recognition Is Accurate, if You're a White Guy

By Steve Lohr

Feb. 9, 2018



Facial recognition technology is improving by leaps and bounds. Some commercial software can now tell the gender of a person in a photograph.

When the person in the photo is a white man, the software is right 99 percent of the time.

But the darker the skin, the more errors arise — up to nearly 35 percent for images of darker skinned women, according to a new study that breaks fresh ground by measuring how the technology works on people of different races and gender.

These disparate results, calculated by Joy Buolamwini, a researcher at the M.I.T. Media Lab, show how some of the biases in the real world can seep into artificial intelligence, the computer systems that inform facial recognition.

Color Matters in Computer Vision

Facial recognition algorithms made by Microsoft, IBM and Face++ were more likely to misidentify the gender of black women than white men.

Article Link