Modern Trends in AI/ML

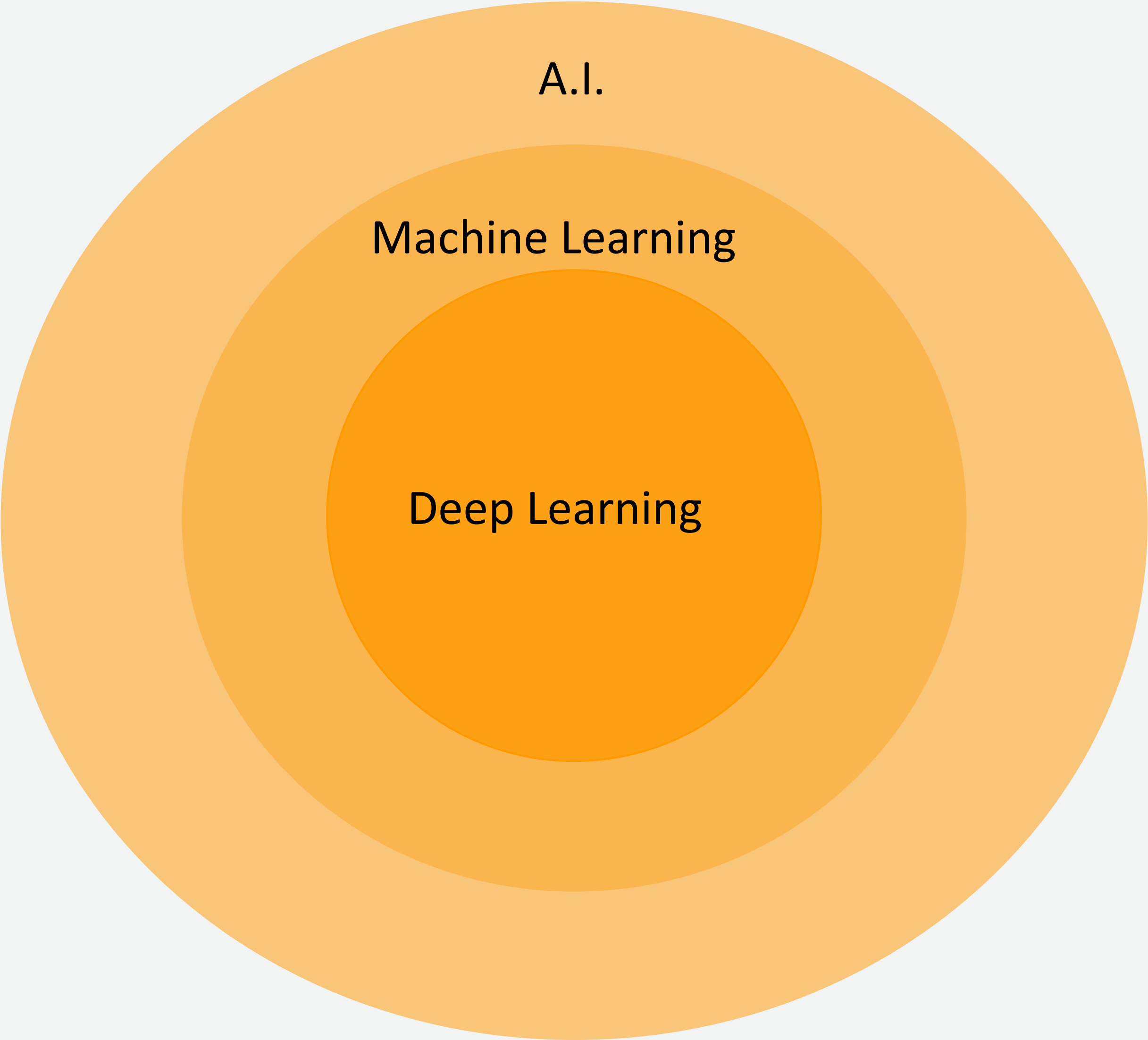
Brad Kenstler
Sr. Data Scientist, Amazon ML Solutions Lab
10 / 24 / 2019

Agenda

1. AI/ML Review
2. Learning from Data
3. Trends in AI/ML application
4. Case Studies

AI/ML Lifecycle Review

AI vs. ML vs. DL



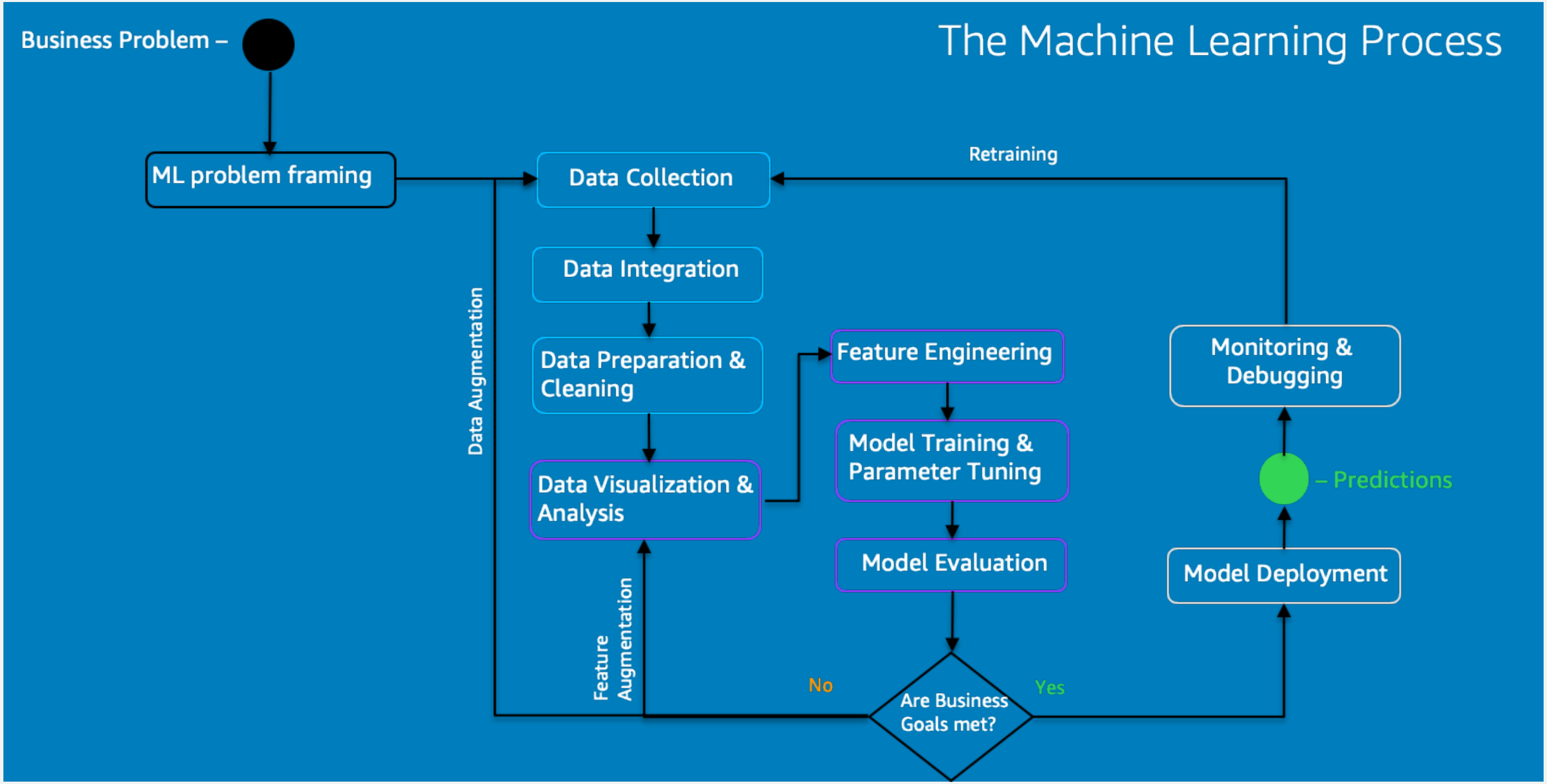
Rules vs. Learning

Rules-based system

- Hand-crafted logic
- Designed by humans to accomplish some goal
- Requires a priori knowledge

Machine Learning

- Learn rules from data that minimize some objective
- May or may not be interpretable
- Requires examples, or **data**



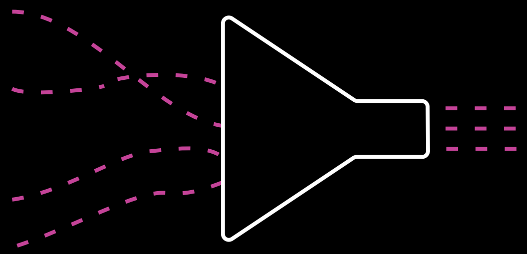
The Machine Learning Process
Business Problem –
ML problem framing
Data Collection
Retraining
Data Integration
Data Preparation & Cleaning
Data Augmentation
Data Visualization & Analysis
Feature Engineering
Model Training & Parameter Tuning
Model Evaluation
Monitoring & Debugging
– Predictions
Model Deployment
Feature Augmentation
No
Are Business Goals met?
Yes

Learning from Data

How do we learn from data?

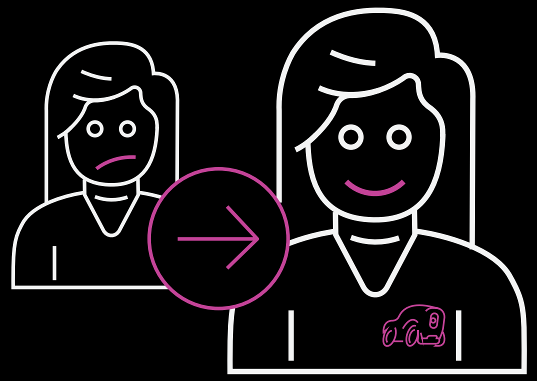
SUPERVISED

Example Driven Training – every datum has a corresponding label



UNSUPERVISED

No labels for training data



REINFORCEMENT

Learns through consequences of actions in a specific environment

Structured Data

Observations or Features

Sepal Length	Sepal Width	Petal Length	Petal Width
0.3	1.2	1.2	2.0
0.5	0.9	1.1	0.9
1.3	1.4	0.9	1.2

Annotation or Labels/Targets

Class
Iris-Setosa
Iris-Virginica
Iris-Versicolour

Unstructured Data

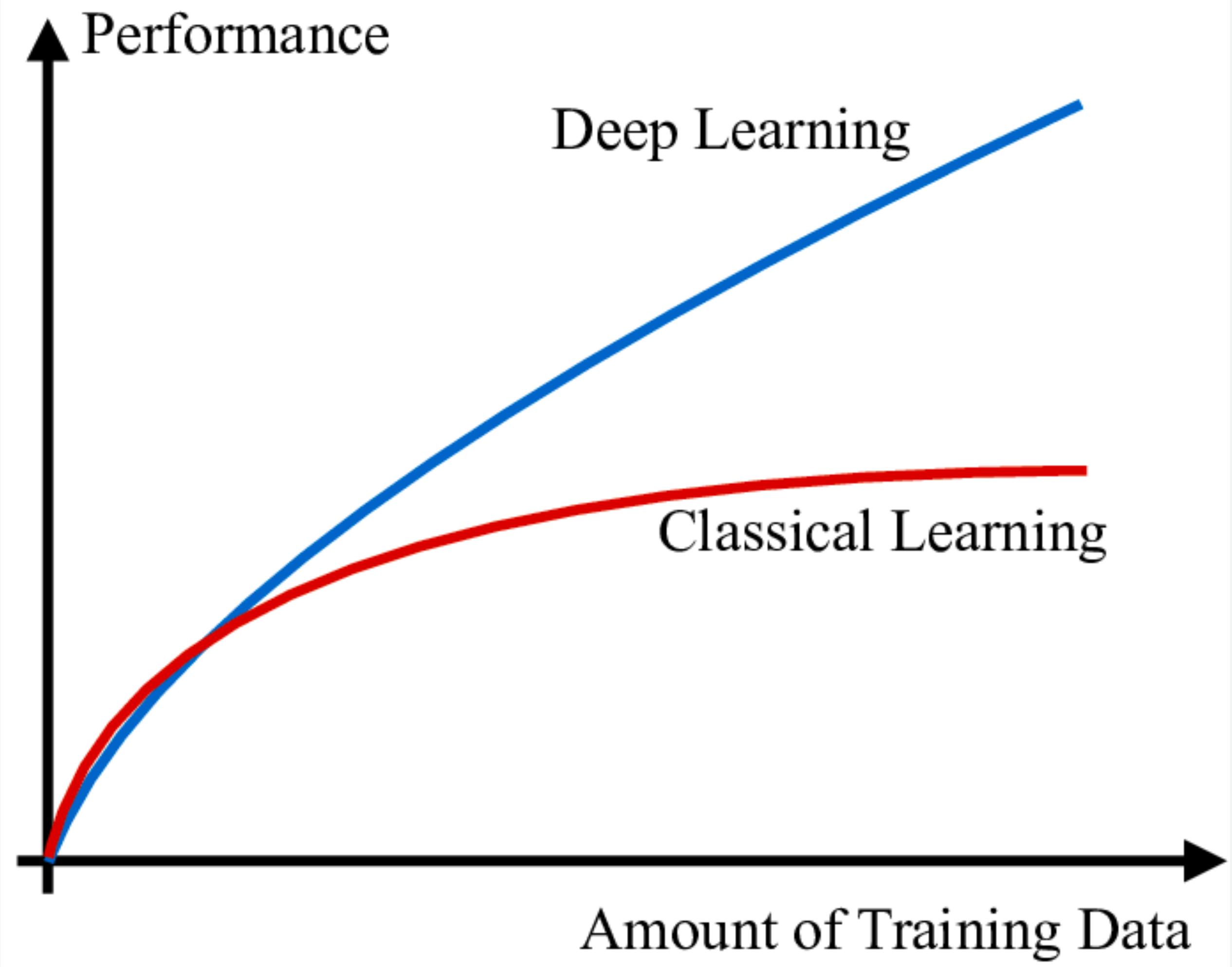
Observations or Features

Annotation or Labels/Targets

How Much Data, and What Kind?

- ML/DL models need large datasets to learn from
- Datasets should be **representative** of the data they are intended to operate on
- Datasets should be large enough to capture expected variation
- Actual amounts depend on use case, but **more is always better**

More Data, Better Results

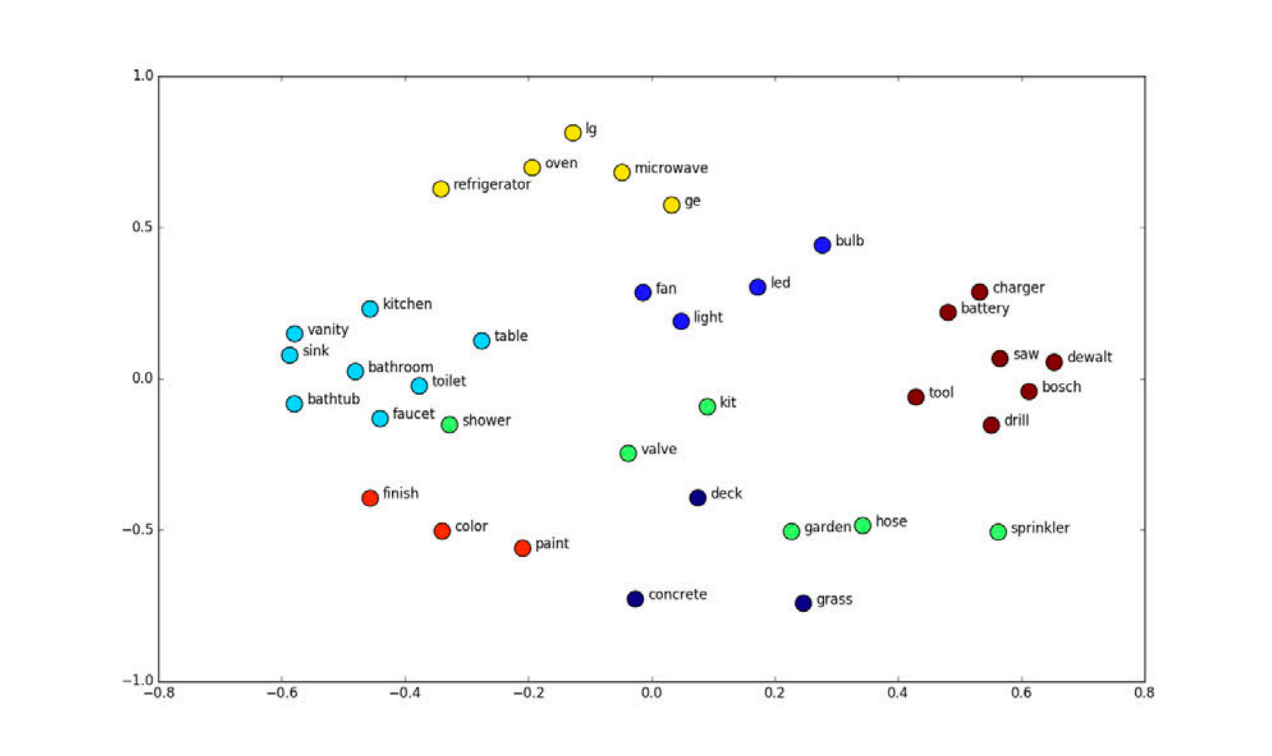


Computer Vision

- Tens of thousands of examples per class
- Hundreds to thousands if using pre-trained models
- Variations in breed, lighting, perspective, setting, etc.

Natural Language Processing (NLP)

- Tens of billions of words in corpus
- Substantially less if using pre-trained models
- Variations in texts, word expression, semantics



Structured Data

Tables

- Thousands to millions
- Representative distribution of features

Time Series

- Hundreds to thousands
- Representative distribution of features over multiple periods

Trends in AI/ML Application

Two Common Scenarios

General-purpose use cases

- ML models trained on millions of data points to do a common task
    - E.g. facial recognition, language translation, voice transcription, etc.
- Completely managed black boxes
- Leveraged via API calls, available as services

Custom use cases

- Custom models trained on custom data set
    - Unique use case or data, e.g. legal document classification
- End-to-end process facilitated by user (with help)

THE AWS ML STACK

Broadest and deepest set of capabilities

AI Services

VISION			SPEECH		LANGUAGE		CHATBOTS	FORECASTING	RECOMMENDATIONS
REKOGNITION IMAGE	REKOGNITION VIDEO	TEXTRACT	POLLY	TRANSCRIBE	TRANSLATE	COMPREHEND & COMPREHEND MEDICAL	LEX	FORECAST	PERSONALIZE

ML Services

Amazon SageMaker

Ground Truth	Notebooks	Algorithms + Marketplace	Reinforcement Learning	Training	Optimization	Deployment	Hosting

ML Frameworks + Infrastructure

FRAMEWORKS	INTERFACES
TensorFlow, mxnet, PYTORCH	GLUON, Keras

INFRASTRUCTURE								
EC2 P3 & P3DN	EC2 G4 EC2 C5	FPGAS	DL CONTAINERS & AMIs	ELASTIC CONTAINER SERVICE	ELASTIC KUBERNETES SERVICE	GREENGRASS	ELASTIC INFERENCE	INFERENTIA

Case Studies

Financial Services Customer ML Solution

Business Issue

Automatic Due Diligence

Support and automate the due diligence process on financial documentation through deep natural language processing

Data

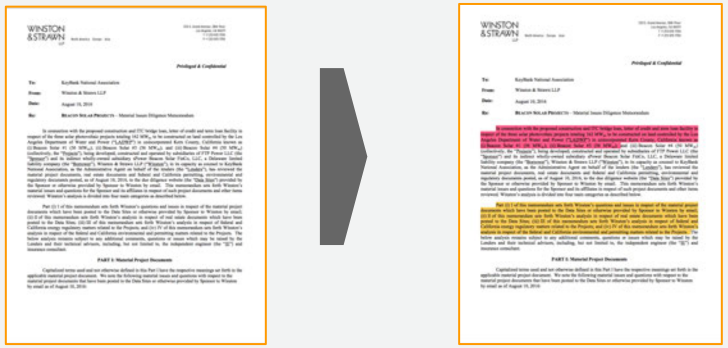
- Corpus of ~1000 documents similar to the target documents (unlabeled)
- Domain specific Wikipedia data
- 2,500 annotated pages (sections of interest highlighted using tagtog, an APN solution)

Machine Learning approach

- Word embedding generation w/ Amazon SageMaker BlazingText
- Recurrent neural network sentence classifier custom built on Amazon SageMaker
- SageMaker integrated frontend based on tagtog to serve predictions natively in pdf and allow users to provide feedback

Outcomes

- **Reducing document due diligence process by 50% (from 8h to 4h)**
- End-to-end natural language processing (NLP) platform to
  - annotate pdf documents
  - train deep learning NLP models
  - provide feedback on machine generated annotations in an easy-to-use user interface to improve the models over time



Retail Customer ML Solution

➢ Business Issue

Markdown Pricing Optimization

Optimize the timing and percentage of markdown (clearance) event for each item at any given week.

➢ Data

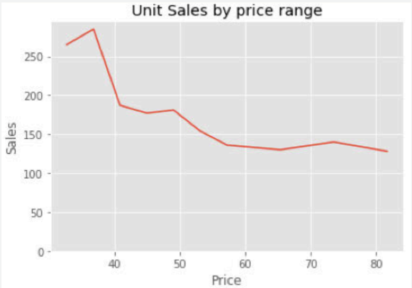
- Time series of weekly item sales and their prices
- Special promotion events
- Item history including the first sale date and base price

➢ Machine Learning approach

- XGBoost and Deep Neural Network for **sales forecasting**
- Customized optimization objective function to optimize a balanced combination of future sell-through rate and profit

➢ Outcomes

- **Optimized Markdown timing and pricing**
- Predicted sell-through rate per item based on multiple inputs
- Maintain a weighted balance between sell-through rate and profit. The relative importance depends on business requirement.



Trade-off between sale and profit based on different pricing scenarios

HCLS Customer ML Solution

➢ Business Issue

Kidney Injury Assessment for Preclinical Toxicity and Efficacy Studies using Computer Vision

Assess overall kidney injury, by locating and scoring all glomeruli each to a pre defined injury class

➢ Data

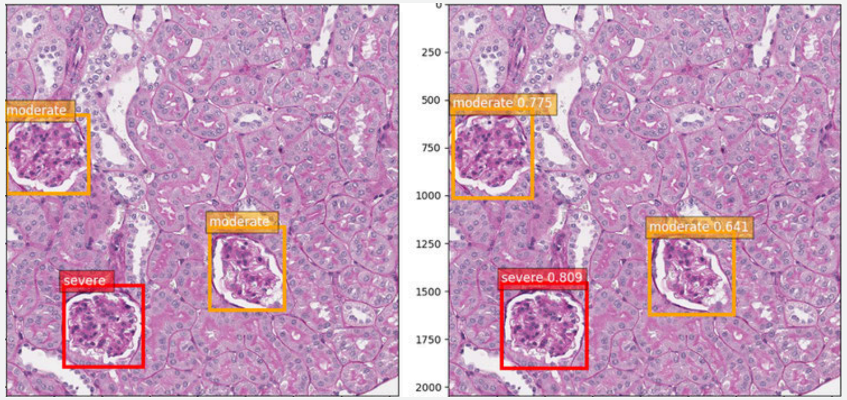
Digital scanned kidney slice images, with annotation of glomeruli including location and injury class information, from customer's diabetic study of mice kidneys.

➢ Machine Learning approach

- **Single Shot Multibox Detection (SSD),** that detects objects in images, and the objects are categorized into one of the classes in the pre defined collection.

➢ Outcomes

- **Detection results were used to enable delivering faster time to scientific insight**
- **Assist manual assessment by increasing accuracy, reducing human errors and bias**
- Reduce operational cost and free up pathologist



Ground Truth — Object Detection

Media & Entertainment Customer ML Solution

➢ Business Issue

Automatic Event and Game Tagging

Identify events in gameplay in **live/near real-time and past games**, to alleviate human needs to log and retrieve key events/replays when prepping for broadcast and analysis

➢ Data

- **Past 3 seasons of player tracking data** from all games
- **Metadata** on key events, players and games

➢ Machine Learning Approach

- Replay retrieval through **linear optimization**, **dynamic time wrapping** and ML **nearest-neighbors**
- Event tagging using combination of multiple **Deep Neural Network** architectures including novel **fully-convolutional networks** and **inception style towers**

➢ Outcomes

- **Used to alleviate human labor as detection of game events in near-real time reached human level accuracy (> 97% overall accuracy across all games)**
- ML models able to **parse tens of thousands of historic plays** and **rank relevant plays** for breakdown and viewing

Manufacturing Customer ML Solution

➢ Business Issue

Automated Visual Inspection

Use **Computer Vision** to augment existing visual inspection during quality control and product grading processes

➢ Data

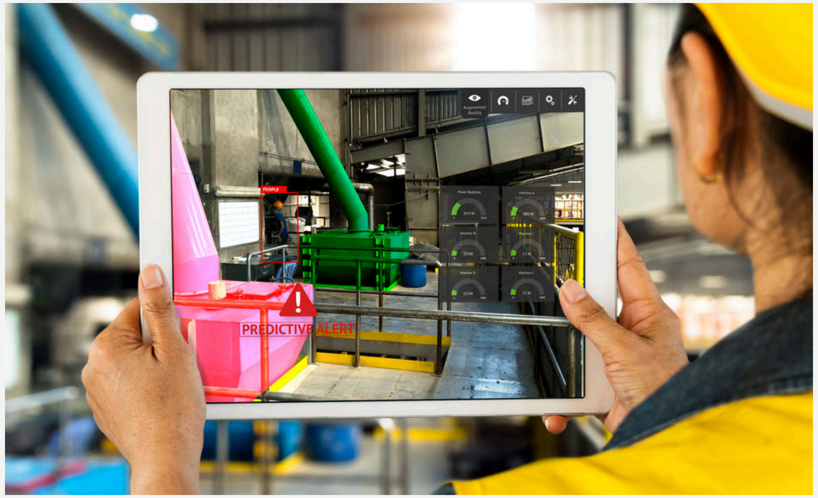
- Visual **images of products** taken during standard inspection process, including RGB and multispectral
- **Failure/non-failure label** across entire set of images OR
- Label on **different types of defects** or different **product grades** across entire set of images

➢ Machine Learning approach

Convolutional Neural Network (CNN) trained on Sagemaker and deployed to **IOT devices**. Leverages **Transfer Learning** with existing pre-trained CNN networks such as ImageNet

➢ Outcomes

- **Results were used to assist manual inspection by increasing accuracy and speed of human inspection for both quality control and product grading**
- Results could also be used to **complement on-going root cause analysis**



Telecom Customer ML Solution

➢ Business Issue

Roaming fraud detection

Detect and prevent roaming fraud in mobile telecommunication network using deep learning

➢ Data

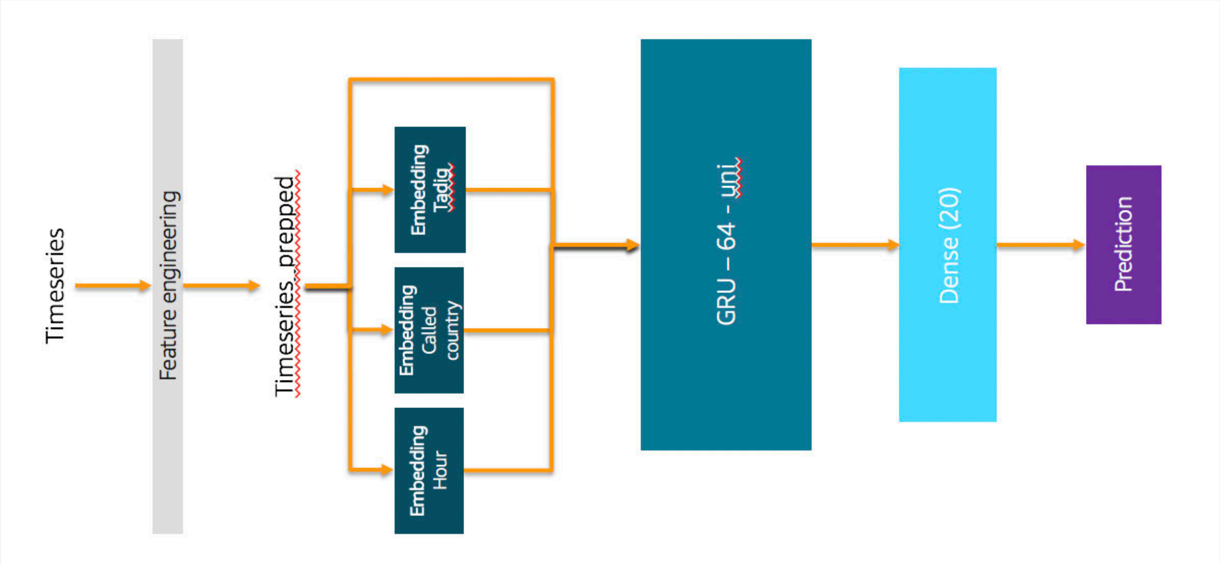
- 350Gb of historical roaming data (standardized across the industry)
- 3500 example cases of identified roaming fraud

➢ Machine Learning approach

- Glue data preprocessing pipeline
- Multi-modal Recurrent Neural Network model

➢ Outcomes

- **Increased the volume of detected fraud by 300% (from 300 to 1200 cases per month)**
- Reduced the time to detection by 80% (from 10h to 2h)



Questions?

Thank you!

Brad Kenstler