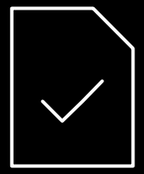
Build, train and deploy Machine Learning models on Amazon SageMaker

Julien Simon
Global Evangelist, AI & Machine Learning
@julsimon

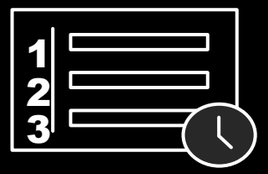
Gàbor Stikkel
Senior Data Scientist
HID Global

aws SUMMIT

Amazon SageMaker

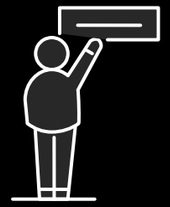


Collect and prepare training data



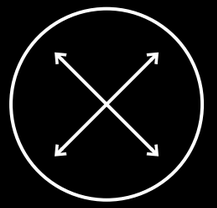
Choose and optimize your ML algorithm

Set up and manage environments for training



Train and Tune ML Models

Deploy models in production



Scale and manage the production environment

Same service and SDK from experimentation to production

intuit

NFL

tinder

KIA

Liberty Mutual.

CONVOY

SIEMENS

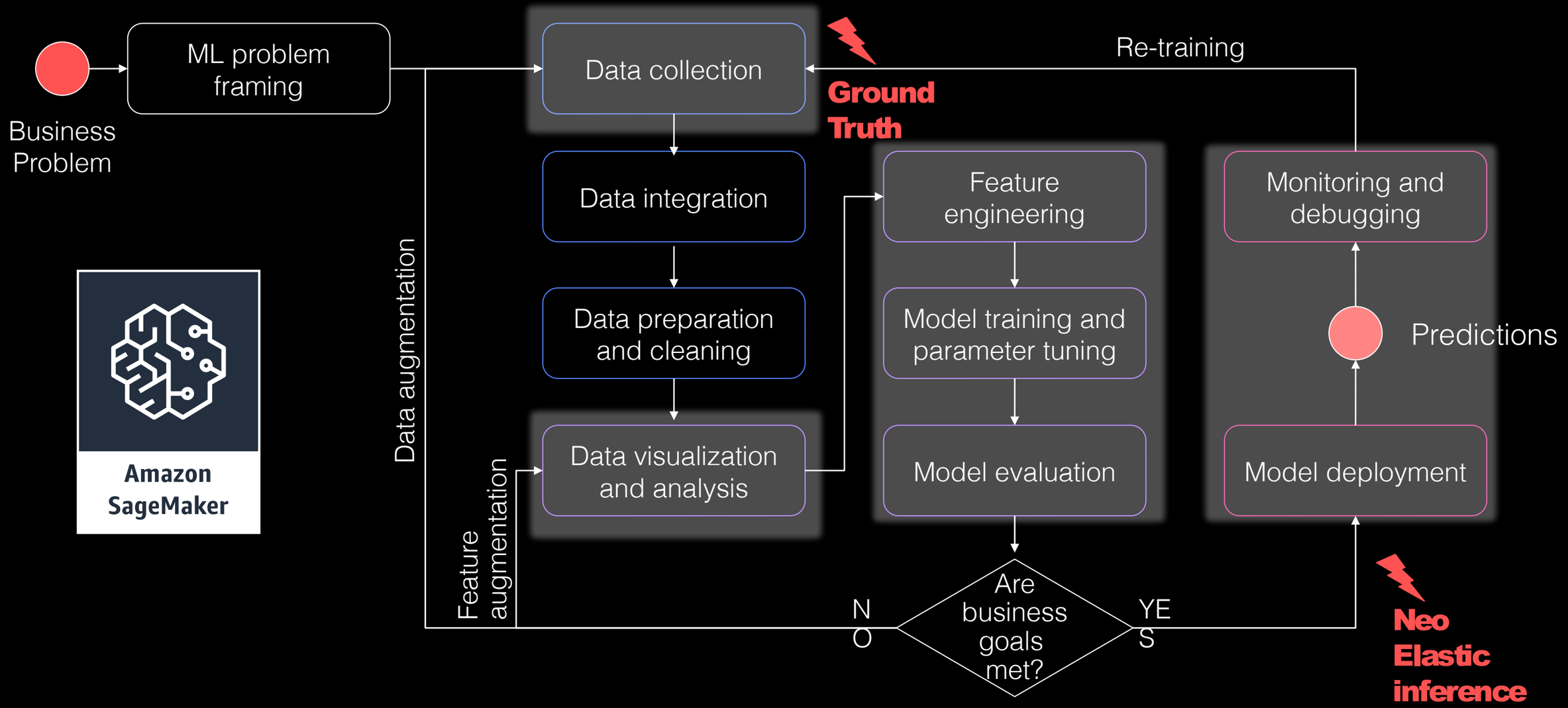
DOW JONES

Celgene

SONY

aws SUMMIT

Build, train and deploy models using SageMaker



The HID Global's Twist Contest

Gesture Recognition in Access Control

AWS Summit Stockholm, 22 May, 2019

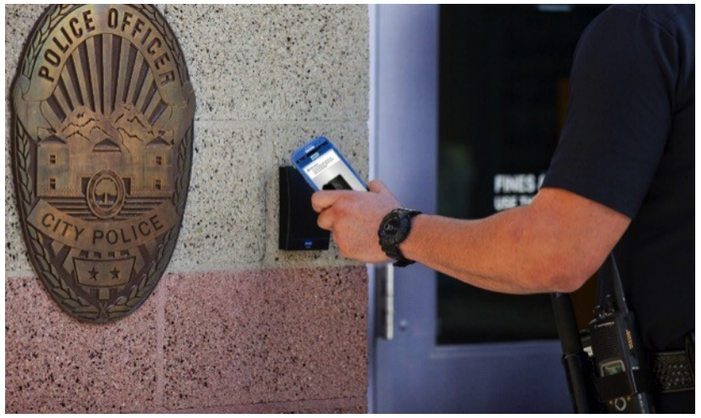
Gábor Stikkel, Senior Data Scientist, HID Global

Every day millions of people in more than
100 countries use our products and services
to securely access physical and digital places

Over 2 billion things
that need to be identified, verified and tracked
are connected through HID's technology

3,200+ employees
worldwide

Part of ASSA ABLOY: 47000+



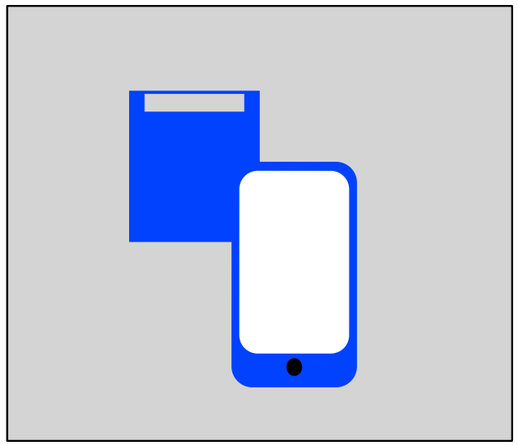
ASSA ABLOY named in
Forbes Top 100 of the
World's Most Innovative Companies

GLASS Trailer (2019)

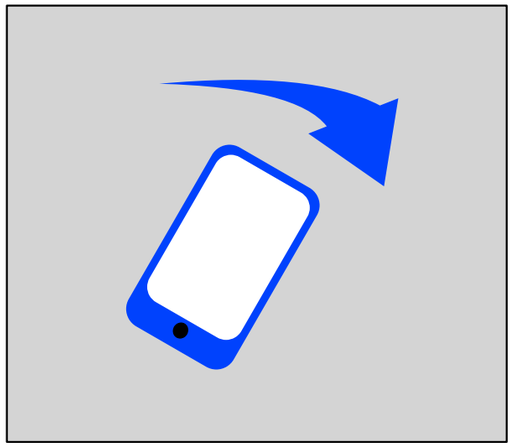
11 003 790 visningar

132 TN 3,4 TN DELA SPARA

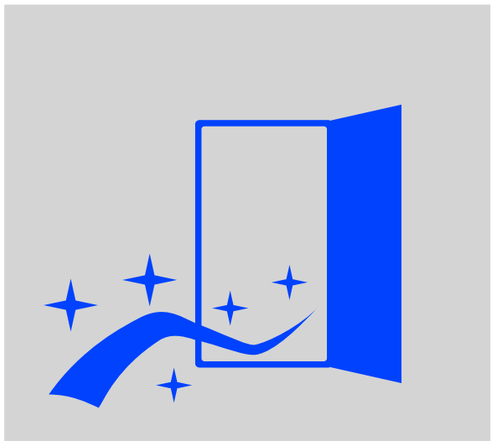
Data-driven use cases in physical access control



Tap

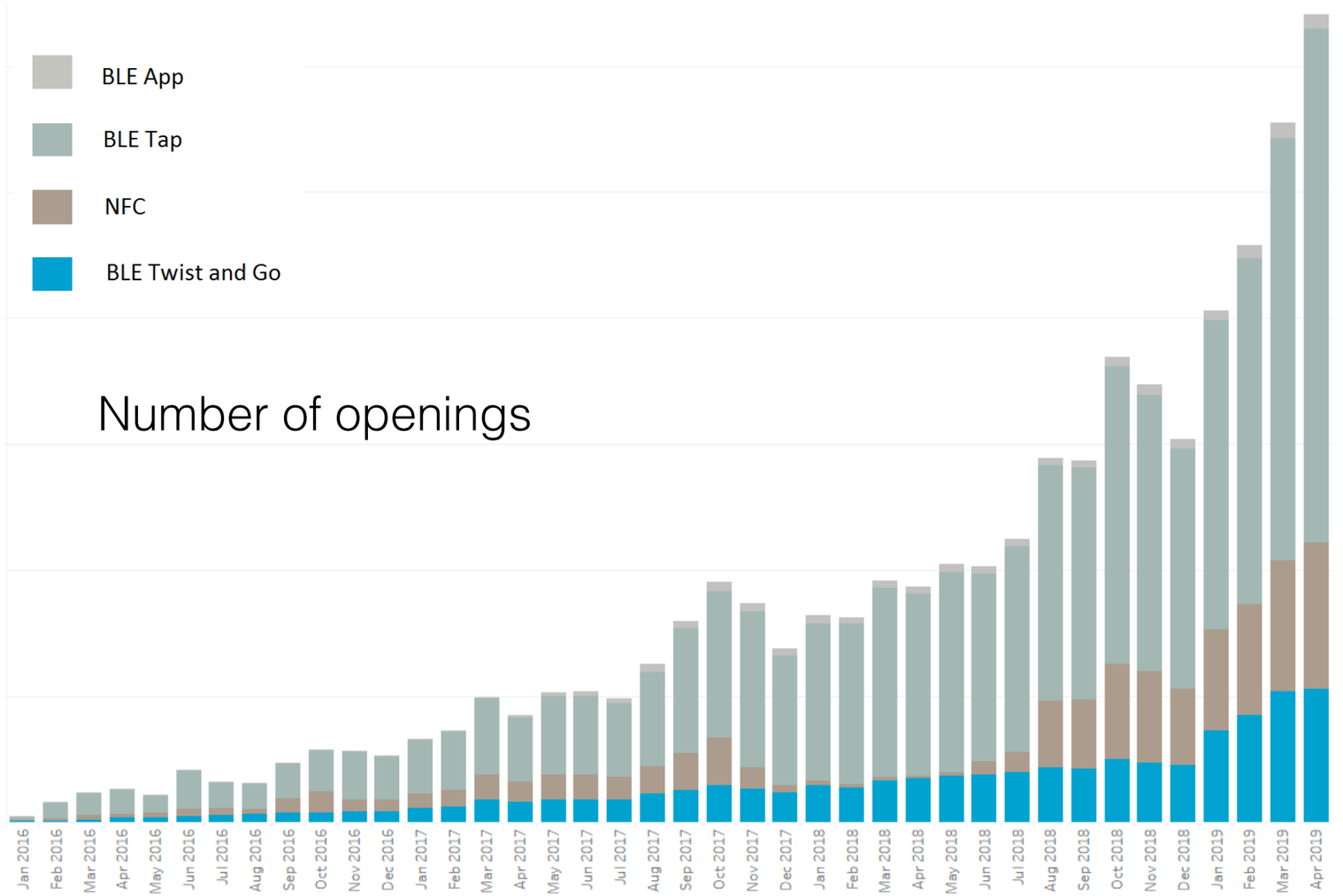


Twist and Go

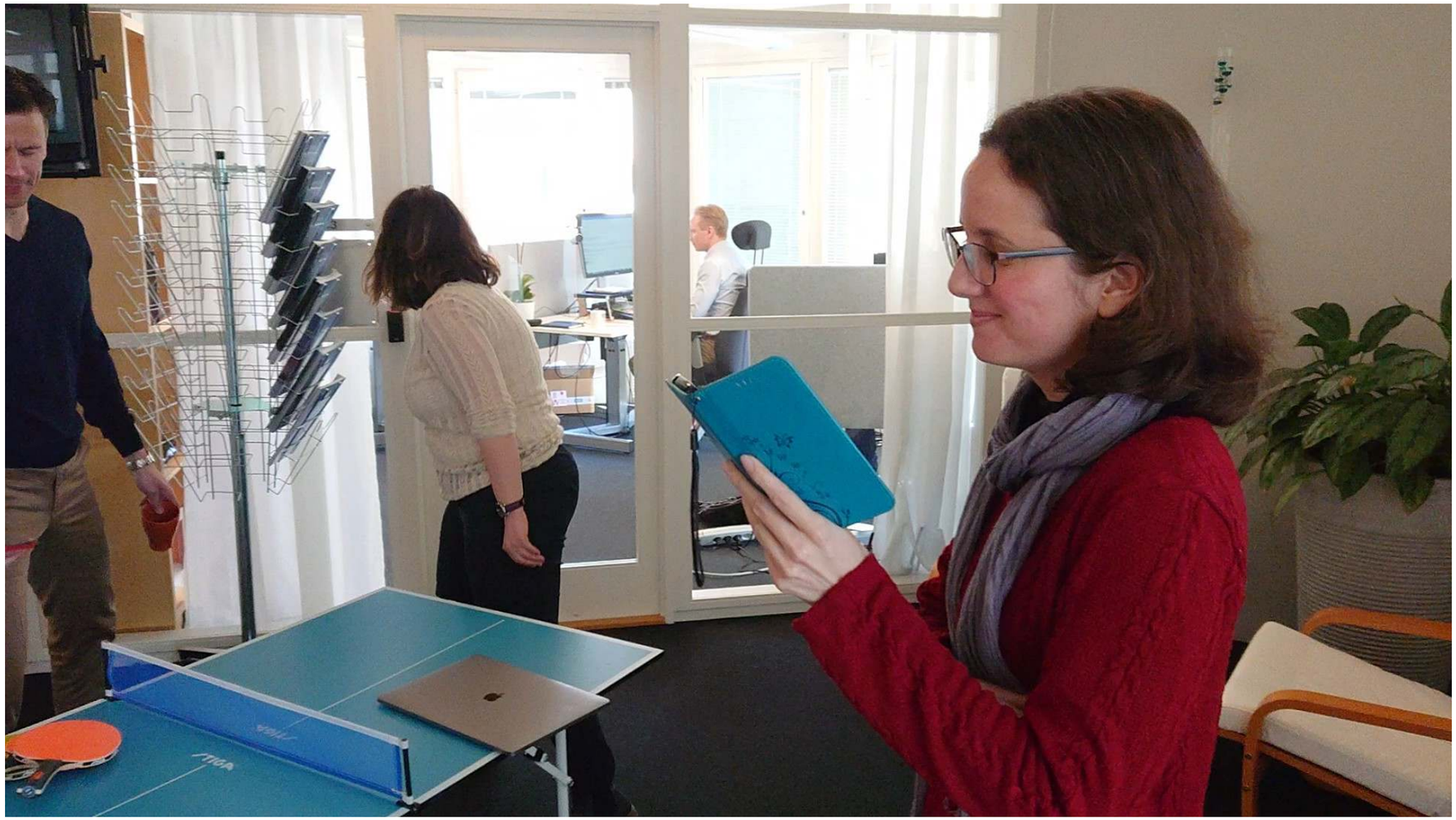


Seamless Access

Launch of Mobile Services was a success



Data collection

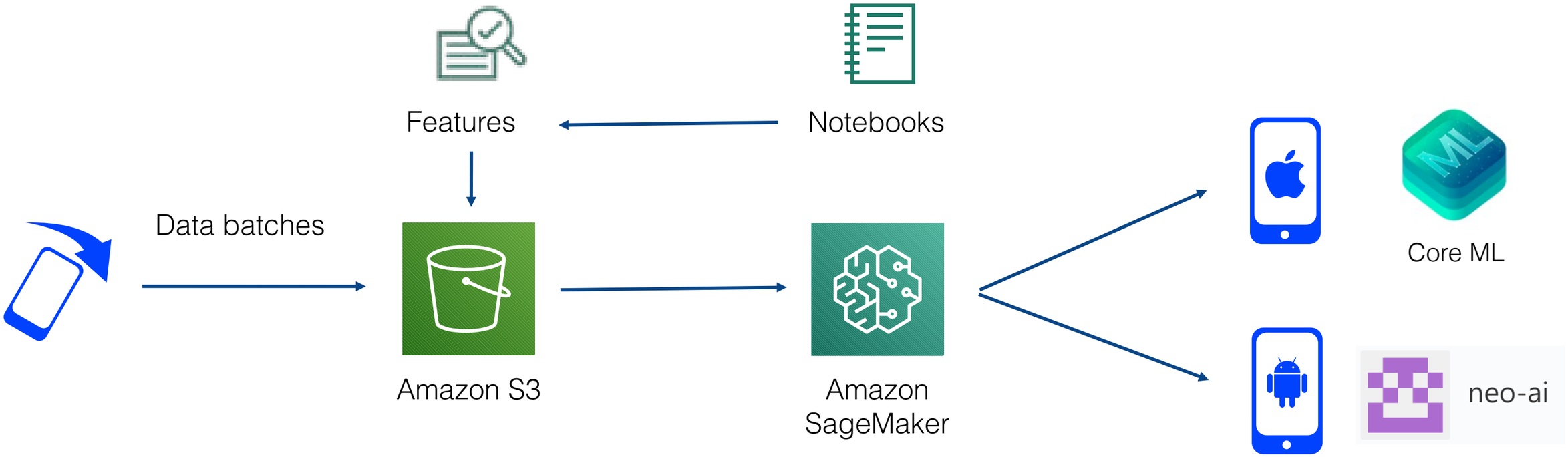


Business problem

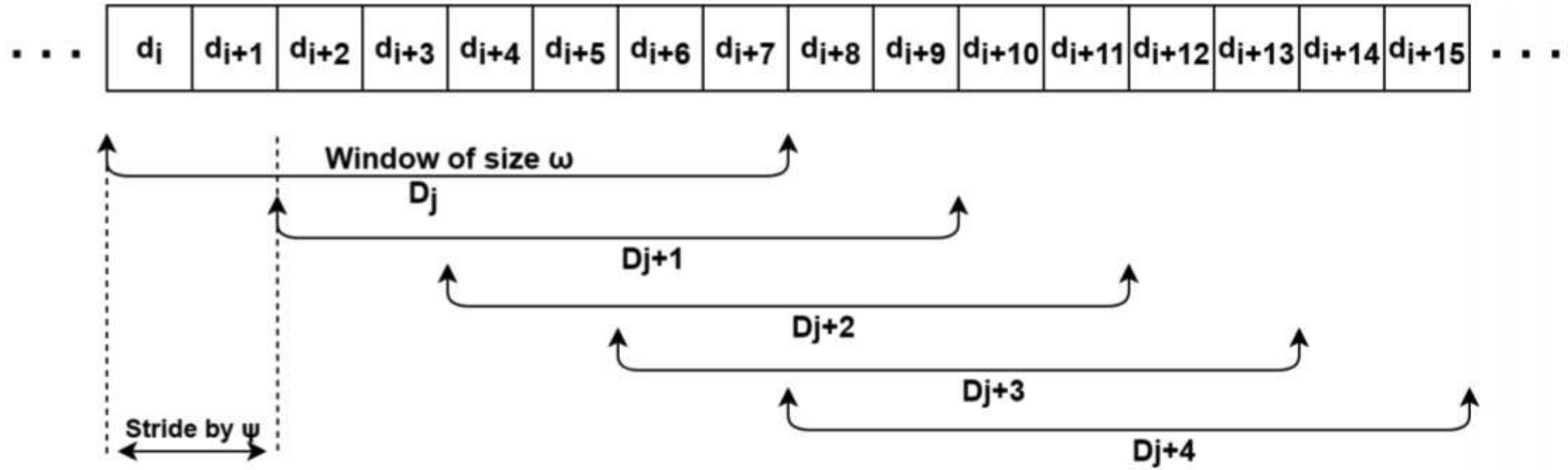
- Simple threshold based rule
- Many different behaviours
- Security issues

Goal: reduce false positives whilst providing a delightful experience

Predictive Modeling Pipeline



Feature calculation using sliding windows



Progammable Gesture Recognition for Augmenting Assistive Devices Sishir Patil at. al. 2018 for details)

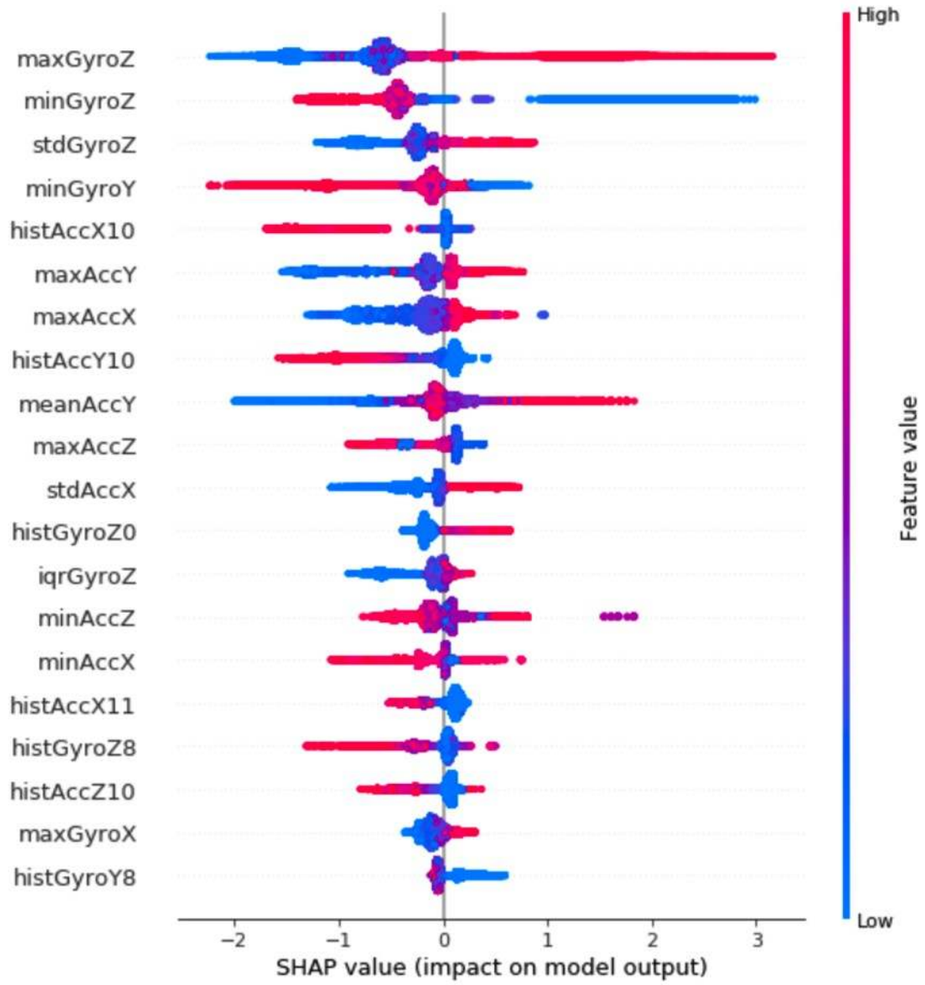
Predictive modeling

Neural networks

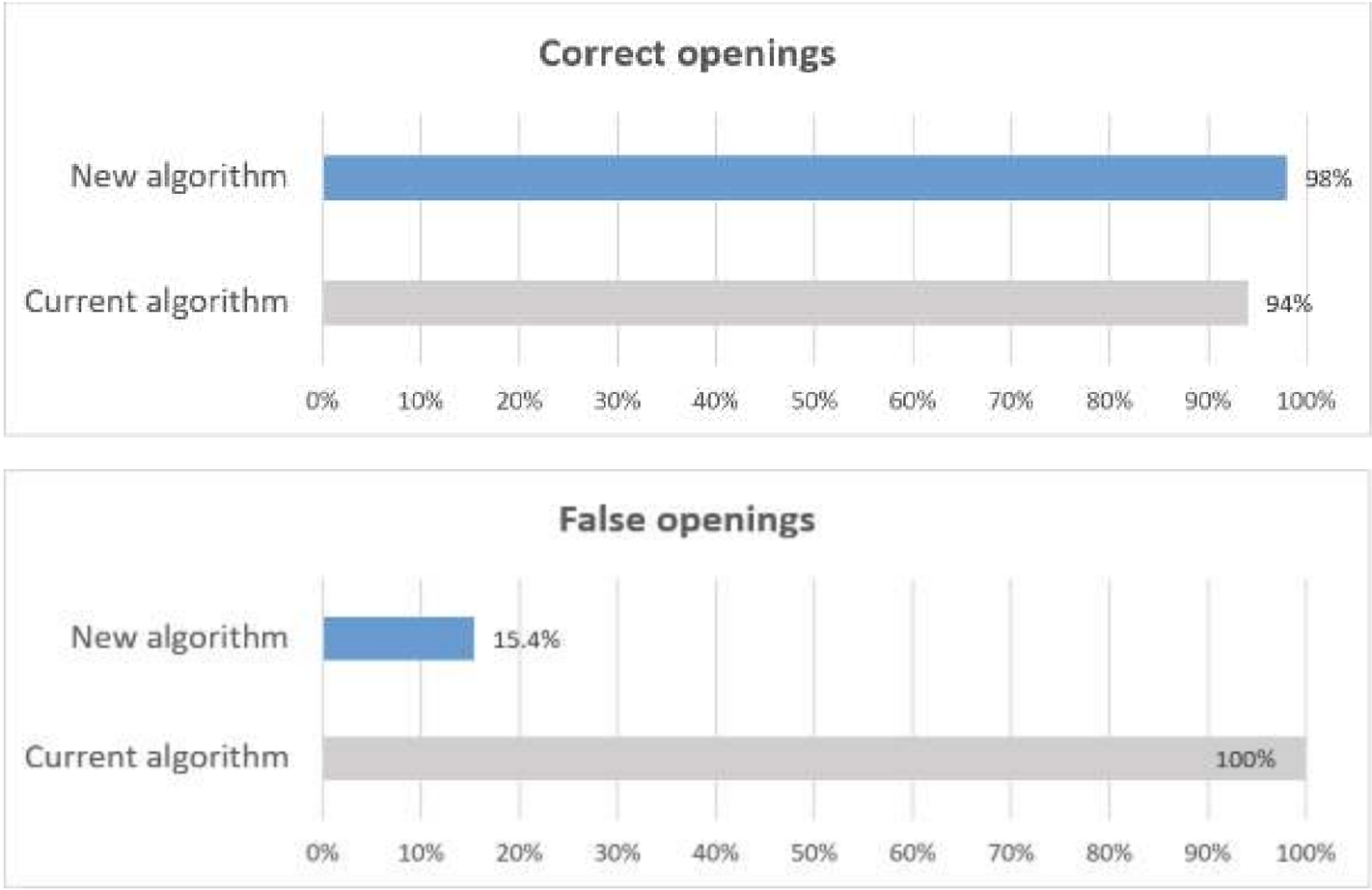
- reproducability issues
- many parameters for even simple models

Tree based ensembles

- better performance
- smaller footprint



Reducing false positives



Extra improvement: twist is recognized ~275ms earlier

?

Phone picture is from https://uae.souq.com/

Tree picture is from https://www.esat.kuleuven.be/

First opening based on a ML model!

2019-01-18 14:23:43.289 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.0010217684

2019-01-18 14:23:43.306 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.002618488

2019-01-18 14:23:43.326 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.0030016804

2019-01-18 14:23:43.345 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.0070577543

2019-01-18 14:23:43.363 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.027276957

2019-01-18 14:23:43.384 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.064488105

2019-01-18 14:23:43.404 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.06967241

2019-01-18 14:23:43.421 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.07492348

2019-01-18 14:23:43.443 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.079435736

2019-01-18 14:23:43.461 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.3511875

2019-01-18 14:23:43.479 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Prediction 0.5662894

2019-01-18 14:23:43.480 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Opening!

2019-01-18 14:23:43.481 17694-17694/com.assaabloy.mobilekeys.android.v2 D/c.a.m.a.e.b.TwistAndGoUltraOpeningTrigger: [main] Twist and Go detected

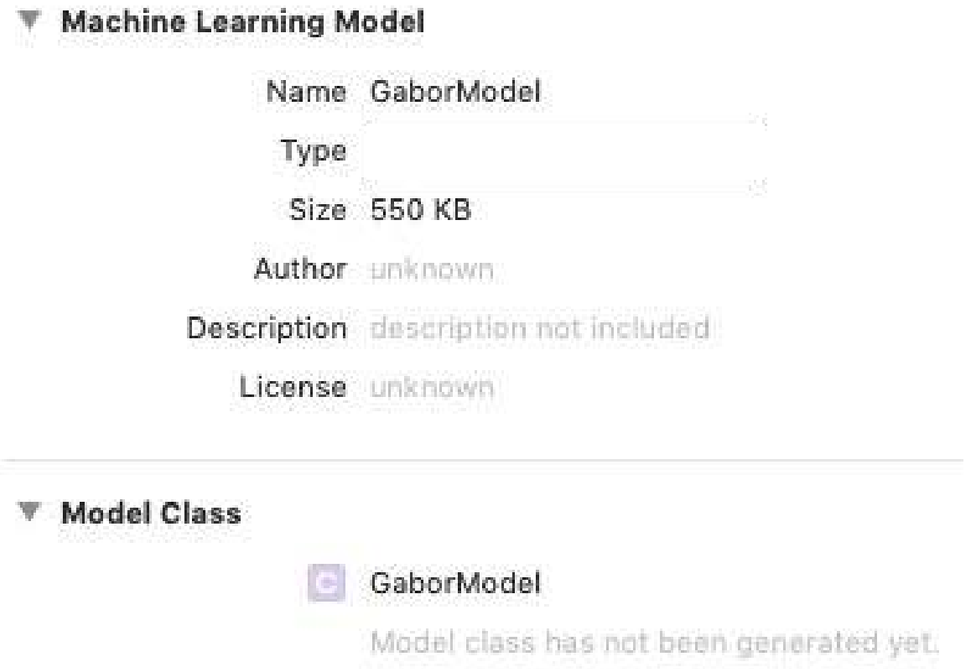
Productification - Android

- treelite to compile the Python model into a C function
- AWS: Neo AI to accelerate model deployment to edge devices
- Keeping feature calculation in synch is hard

```c
#include "header.h"

size_t get_num_output_group(void) {
  return 1;
}

size_t get_num_feature(void) {
  return 174;
}

static inline float pred_transform(float margin) {
  const float alpha = (float)1;
  return 1.0f / (1 + expf(-alpha * margin));
}

float predict(union Entry* data, int pred_margin) {
  float sum = 0.0f;
  unsigned int tmp;
  if (!(data[2].missing != -1) || (data[2].fvalue < 6.3081808)) {
    if (!(data[145].missing != -1) || (data[145].fvalue < -4.9956379)) {
      if (!(data[121].missing != -1) || (data[121].fvalue < -3.893259)) {
        if (!(data[59].missing != -1) || (data[59].fvalue < -5.1344109)) {
          if (!(data[146].missing != -1) || (data[146].fvalue < 0.27744454)) {
            if (!(data[156].missing != -1) || (data[156].fvalue < 6.5)) {
              sum += (float)0.17714286;
            } else {
              sum += (float)-0.022222223;
            }
          } else {
            if (!(data[8].missing != -1) || (data[8].fvalue < 6.449605)) {
              sum += (float)-0.12;
            } else {
              sum += (float)-0;
            }
          }
        } else {
          if (!(data[133].missing != -1) || (data[133].fvalue < 12.5)) {
            if (!(data[46].missing != -1) || (data[46].fvalue < 9.5)) {
              sum += (float)-0.14202899;
            } else {
              sum += (float)0.06666667;
            }
          } else {
            if (!(data[0].missing != -1) || (data[0].fvalue < -11.438884)) {
              sum += (float)-0;
            } else {
              sum += (float)0.1;
            }
          }
        }
      } else {
        if (!(data[75].missing != -1) || (data[75].fvalue < 10.5)) {
          if (!(data[50].missing != -1) || (data[50].fvalue < 16.5)) {
            if (!(data[117].missing != -1) || (data[117].fvalue < -0.98243845)) {
              sum += (float)-0.022222223;
            } else {
              sum += (float)-0.17352942;
            }
          } else {
            if (!(data[90].missing != -1) || (data[90].fvalue < 0.38042349)) {
              sum += (float)-0.13846155;
            } else {
              sum += (float)0.10476191;
            }
          }
        } else {
```

Productification - iOS

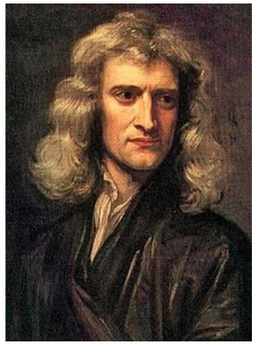
- More black-box than Android
- Same challenges with feature calculation



Conclusions

Data is the new water – it comes from every tap
People are unpredictable – they invent all sorts of gestures
Lowest hanging fruits are grown on decision trees

”ML tool support from AWS making data scientists' life easier”



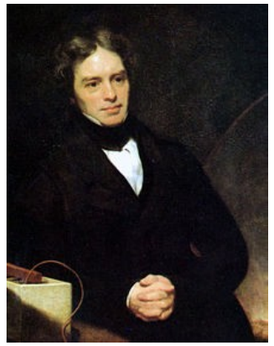
Isaac Newton
1643 - 1727

C-A. de Coulomb
1736 - 1806

Alessandro Volta
1745 - 1827

André-Marie Ampère
1775 - 1836

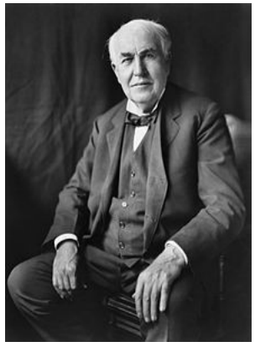
Georg Ohm
1789 - 1854



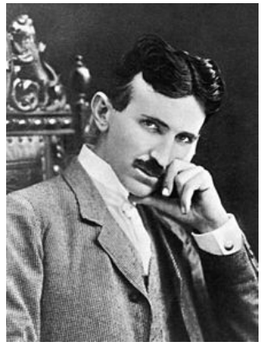
Michael Faraday
1791 - 1867

Joseph Henry
1797 - 1878

Ada Lovelace
1815 - 1852



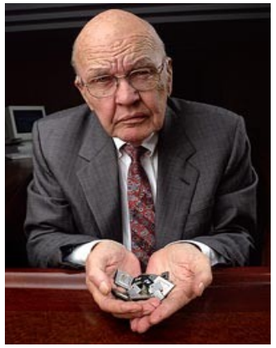
Thomas Edison
1847 - 1931



Nikolas Tesla
1856 - 1943

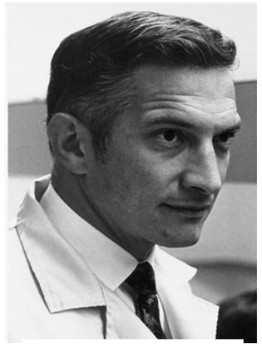
Arthur Samuel
1901 - 1990

János Neumann
1903 - 1957



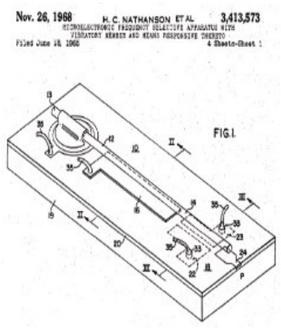
Jack Kilby
1923 - 2005

Transistor
1926 -

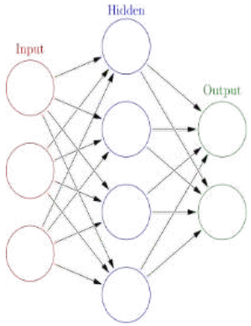


Robert Noyce
1927 - 1990

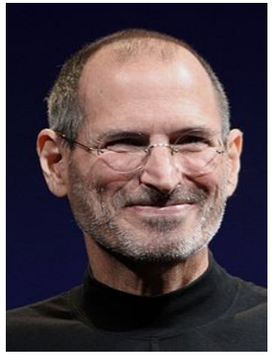
Leo Breiman
1928 - 2005



Harvey Nathanson
1936 -



McCulloch - Pitts
1943



Steve Jobs
1955 - 2011

Tianqi Chen

Pictures imported from wikipedia.org

The five beer team

Deep Learning on Amazon SageMaker

aws SUMMIT

Model options

Training code

AWS Machine Learning Marketplace: 150+ off-the-shelf models

Factorization Machines
Linear Learner
Principal Component Analysis
K-Means Clustering
XGBoost
And more

Built-in Algorithms (17)

No ML coding required
No infrastructure work required
Distributed training
Pipe mode

mxnet Chainer TensorFlow PyTorch scikit learn

Built-in Frameworks

Bring your own code: script mode
Open source containers
No infrastructure work required
Distributed training
Pipe mode

Bring Your Own Container

Full control, run anything!
R, C++, etc.
No infrastructure work required

Built-in Deep Learning frameworks: just add your code

- Built-in containers for training and prediction.
  - Code available on Github, e.g. https://github.com/aws/sagemaker-tensorflow-containers
  - Build them, run them on your own machine, customize them, etc.
- Script mode: use the same code as on your laptop

No infrastructure work required: simply define instance type and instance count

Distributed training out of the box: zero setup

Pipe mode: stream infinitely large datasets directly from Amazon S3

AWS: The platform of choice to run TensorFlow

85% of all TensorFlow workloads in the cloud runs on AWS

Source: Nucleus Research, November 2018

Optimizing Tensorflow for Amazon EC2 instances

C5 instances (Intel Skylake)

Training ResNet-50 with the ImageNet dataset using our optimized build of Tensorflow 1.11 on a c5.18xlarge instance type is 11x faster than training on the stock binaries.

P3 instances (NVIDIA V100)

Tensorflow scaling efficiency with 256 GPUs

65 % Stock version

90 % AWS-optimized version

Apache MXNet: Deep Learning for enterprise developers

Start with off-the-shelf models

- Gluon CV and Gluon NLP
- ONNX compatibility

Fast and scalable training

- Keras-MXNet up to 2x faster than Keras-TensorFlow
- Near-linear scalability up to 256 GPUs
- Dynamic training

Easy deployment

- Java/Scala APIs
- Model Server

Demo: Keras+Tensorflow

Script mode

Automatic model tuning

Elastic inference

https://gitlab.com/juliensimon/dlnotebooks/tree/master/keras/04-fashion-mnist-sagemaker-advanced

Getting started

http://aws.amazon.com/free

https://aws.amazon.com/sagemaker

https://github.com/aws/sagemaker-python-sdk
https://github.com/awslabs/amazon-sagemaker-examples

https://medium.com/@julsimon
https://gitlab.com/juliensimon/dlnotebooks

Thank you!

Julien Simon
Global Evangelist, AI & Machine Learning
@julsimon

Gàbor Stikkel
Senior Data Scientist
HID Global

aws SUMMIT

Please complete the session survey.

aws SUMMIT