



# Decision Intelligence & Streaming Predictive Analytics with Spark

Scott Haines  
Senior Principal Software Engineer  
Voice Insights

<http://bit.ly/odsc-2020-streaming-di>



# Table of contents

---

- 1 Introductions
- 2 Decision Making Process
- 3 Data Pipelines & Streams
- 4 Semi-Autonomous and Autonomous Decisioning
- 5 Hands-On Workshop

<http://bit.ly/odsc-2020-streaming-di>



# Introductions

---





# Scott Haines

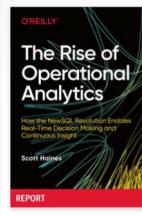
Senior Principal Engineer  
Voice Insights, Twilio

---



Over Ten Years working on  
Streaming Systems

Insights, Discovery,  
Recommendations, Personalization,  
Fraud, Trends, Ads, and Mobile  
Campaigns



Recent Release

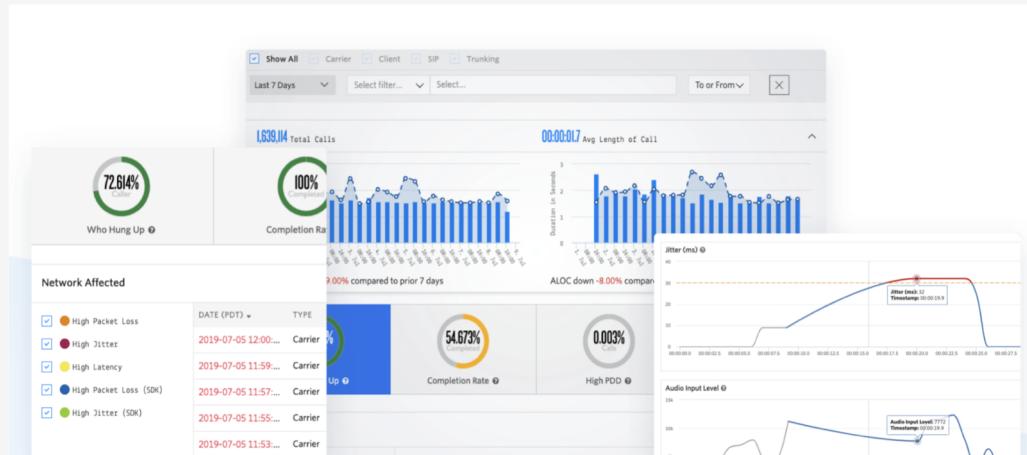
[The Rise of Operational Analytics](#),  
O'Reilly



# Voice Insights

Interactive Operational Analytics-as-a-Service for Voice.

Ridiculous Scale. Powered by Spark



[Learn More](#)



## Deliver exceptional customer experiences

Interact with your customers on the channels they prefer at every step of their journey across marketing, operations, and customer service.



Try a phone call from the Voice API

415-555-5555

Demo available in the U.S. and Canada.

Call my phone

PHP NODE PYTHON RUBY JAVA .NET (C#) CLI CURL

```
# Download the helper library from https://www.twilio.com/docs/python/install
from twilio.rest import Client

# Your Account Sid and Auth Token from twilio.com/console
account_sid = 'ACXXXXXXXXXXXXXXXXXXXXXX'
auth_token = 'your_auth_token'
client = Client(account_sid, auth_token)

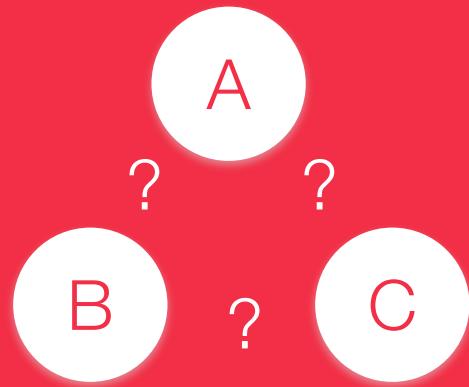
call = client.calls.create(
    url='http://demo.twilio.com/docs/voice.xml',
    to='+15558675310',
    from_='+15017122661'
)
print(call.sid)
```

Customer Engagement Cloud



# Decision Making Process

---



How Do We  
Teach Machines  
To Think...



# Decision Making Process

---

If We Can't  
Define What We  
Want Them To  
Know?



## Decisions are Tough

Do you wake up? Snooze? Turn Left or Right? *Avoid that street* or Take a known short cut?

---

### Decision Indecision

When faced with one or more difficult decisions, What do you do?

### Fight or Flight

Instinctive decision making is rooted in nature and based on *encoded historic observations*.

### Stay the Course or Explore

Do you stick to your what you know ("your gut") or do you explore and find new paths that may yield better reward?



# Decision Maker

Machine Learning projects need someone appointed to the role of the Decision Maker.

---

What is missing that is critical to the Business or Project?

The decision maker is a conduit to the bigger picture. Why are we investing in X or Y?

How do we define a Data Story?

What Questions do we *need to* ask of our Data? Do we have Data to answer these questions? How do we know it is valid and honest?

How do we define a Decision Domain?

Can we create a general framework for decisioning within a problem domain?



# Decision Maker

Thinks about *the expectations of the data* for the Business or Project.

---

## Identify a Use Case or Blind Spot

We need to understand why our customers keep cancelling our service.

## Build up a Story

By understanding *the typical behavior* of our customers, we should be able to reduce churn through personalization and other incentives.



# Decision Making

What does your Data Tell You? How can you measure the success of your decisioning?

towards  
DATA SCIENCE PROGRAMMING VISUALIZATION

Storytelling with Data  
The subtle art of good analytics and explainable insights

Scott Haines Oct 8, 2019 - 6 min read

Storytelling with Data isn't as whimsical. It does, however, provide the correct mental model. (Source: [Flowtop](#))

the article

## Break into Scenarios

Does the cost of the service influence retention?

Do customers churn at specific times of year?

Do customers in specific regions churn more often than others?

What other dimensions or segments can we identify...



# What Is Your Data Telling You?

Think about *the expectations of the data* you collect. What surface area does the data cover? Is it raw or pre-processed?

The subtle art of good analytics and explainable insights

Scott Haines  
Oct 8, 2019 - 6 min read

Storytelling with Data isn't as whimsical. It does, however, provide the correct mental model. (Source: [Plotly](#))

the article

## Identify Sources of Data

Do we have available data to assist in our decision making?

Can we use the data? eg. Have customers opted-in? GDPR?

Can we trust the data?

What is the lineage of the data?

Is there a *Data Steward* responsible for quality of the data?

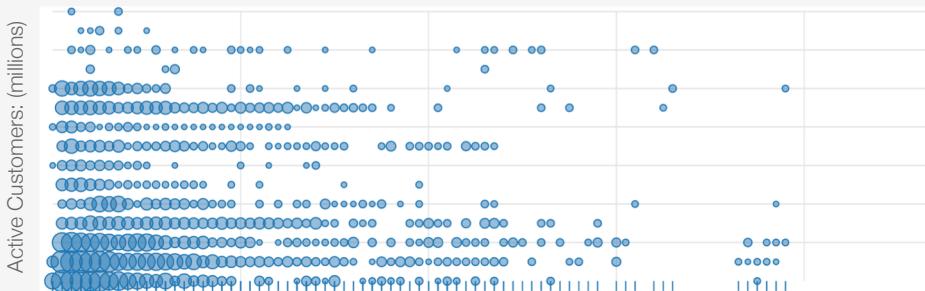
Do we need to start from scratch?

Is there an Open-Source dataset or can we purchase datasets?



# What Is Your Data Telling You?

Data Analysts / Data Detectives / Explorers



Unique Users Region Per Day (last 6 months)

Explore / Identify / Repeat

For each scenario, within each story, can we identify patterns to backup our assumptions?

Do we need to go back to the drawing board?

Data at this stage can be raw, semi-processed, or fully processed. If this is a new investigation or theory then notebook environments like [Zeppelin](#) and [Jupyter](#) help teams to share results across the board.



# Feature Engineering

Take human readable data and prepare it for Machine Learning models. Remember we are not machines and need more context!

---

## Prepare

Take human readable data and prepare it for Machine Learning models

## Scale

Identify and Reshape data distributions to fit normal distributions or normalize the underlying data

## Index / Encode

Take categorical values and class labels and convert these string representations into encoded vectors. StringIndex / OneHot Encoding



# Model Selection / Model Tuning

Given a clean, prepared dataset, tackle the model selection and tuning problem.

---

## Type of Data

Do you have labels? Supervised otherwise  
Unsupervised.

## Classification or Regression?

Identify and Reshape data distributions to fit normal distributions or normalize the underlying data

## Save for Later

Export your trained model and the preserved Test dataset so you can continue to test enhancements.



# Prediction to Decision

Machine Learning Models predict an **outcome** given an **input**. Taking an **action** based on the prediction is a **decision**.

---

## Fraud Detection

Historic Fraud patterns can predict what is *occurring now*, or forecast probability of future fraud. Action is to block fraudulent behavior.

## Growth and Re Engagement

Customers who follow known patterns of disengagement can be isolated for outreach based on statistical machine learning model.

## Cost-Risk Optimization

Farming is beginning to embrace Machine Learning for smart watering. Leveraging live rain forecasts and smart IoT sensors. Take actions to produce better crops.

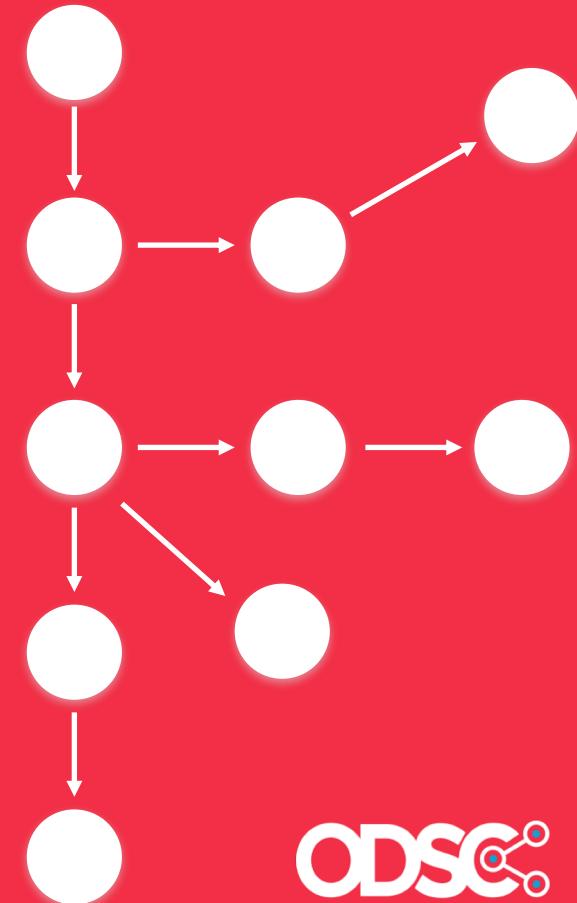


Decision Intelligence holds humans  
accountable for automated decisioning



# Data Pipelines & Streams

---





# Chicken & The Egg

Data Engineers can help to add strict validations and schema enforcement prior to ingestion into your data streams.

---

## Model the Data

Streaming systems need something to process. It is typical to model data into the top level domains of Events and Metrics.

## Prepare your Data for Streaming

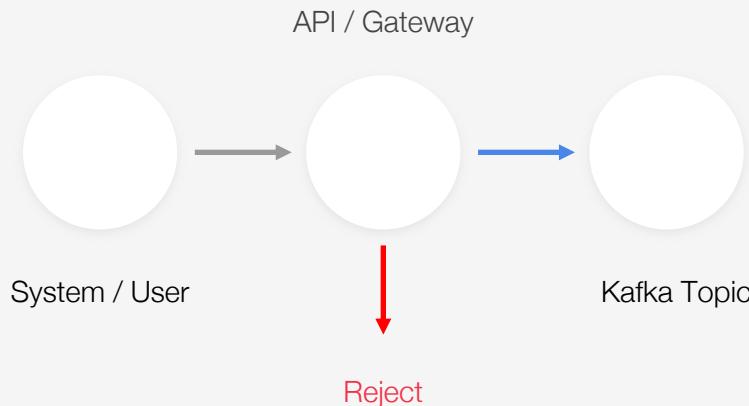
Streaming systems are always online. 24/7/365. Corrupt data can bring your entire pipeline to a grinding halt.

## Strictly Enforced Schemas

Use a Structured Data library like [Avro](#) or [Protobuf](#) to encode your data into versioned, language agnostic formats.

# Enforce Early

Unless you are capturing raw data for experimentation. *Enforce and Encode* your data before putting it into a Stream.



## Validate at the API or Gateway

Searching down corrupt data is hard enough without tracing it all the way to the root.

## Encode at the Service

Stand up an API in front of your Data Pipelines or use a Client Library / SDK to enforce data feeding into your system. This means all data can be uniformly encoded into the current version of your Data Models.



# Semi-Autonomous & Autonomous Decisioning

---

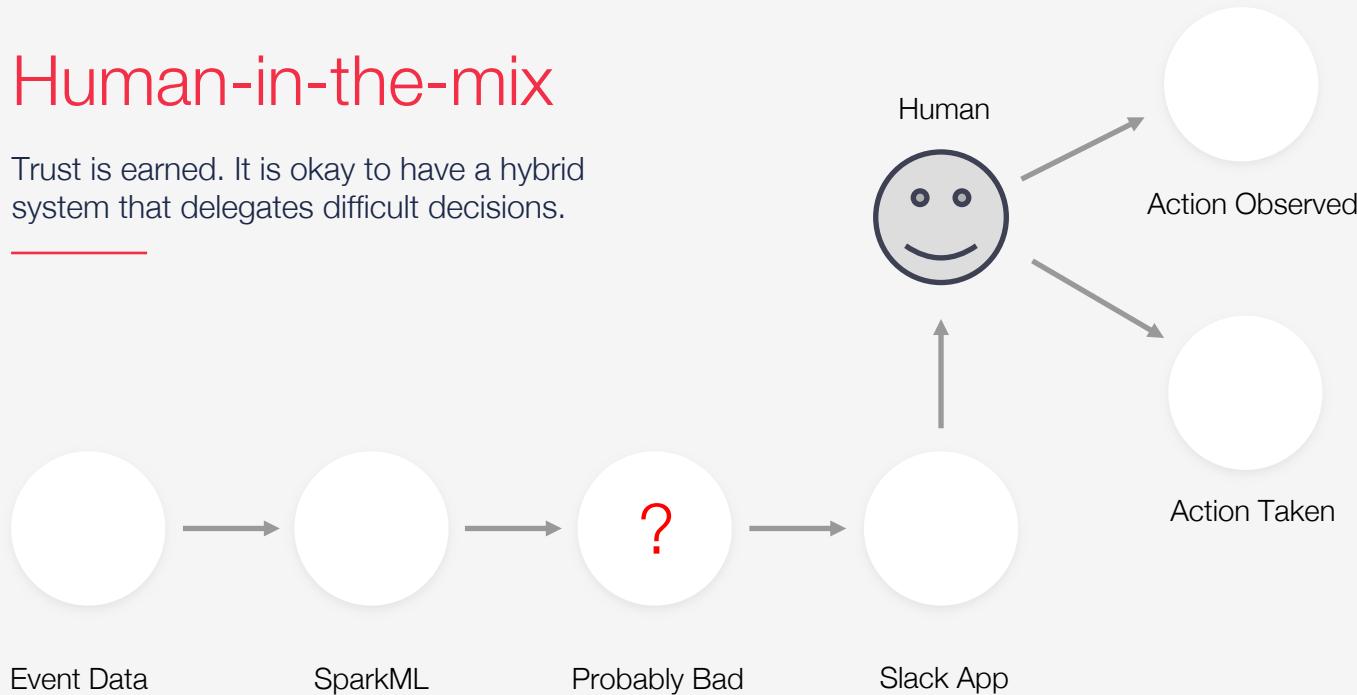




# Human-in-the-mix

Trust is earned. It is okay to have a hybrid system that delegates difficult decisions.

---





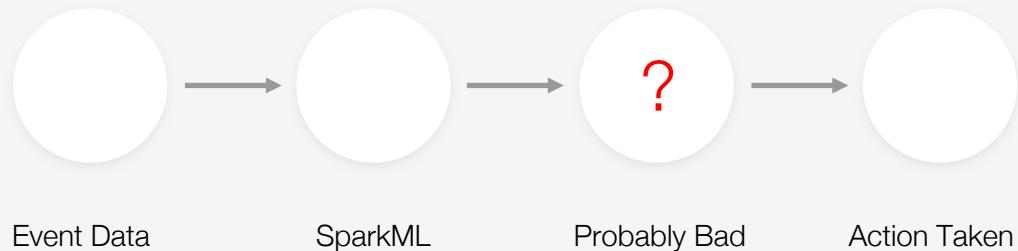
# Fully Autonomous

After trust is earned. Decision making can be delegated to your system.

---

## Rinse and Repeat

Risky changes to your models can continue to run through semi-human-supervised mode.



# Workshop

---

# Building Smart Apache Spark Structured Streaming Applications

<http://bit.ly/odsc-2020-streaming-di>



Thank you