

# MACHINE LEARNING IN PPC

## How To Get Started Today

Christopher Gutknecht | norisk Group | #FOS19



# Our Agenda: Intro & 3 PPC Use Cases

INTRO

Machine Learning Essentials

ML Toolkit For PPC

Platforms & Tools

Query Understanding

1. Classifying Near-Exact

Prediction

2. Analysing Query n-Grams

Text Summarization

3. Finding Key Phrases

# Note: I'm not a Data Scientist - I'm a PPC



Christopher Gutknecht

Head of Online Marketing @ norisk

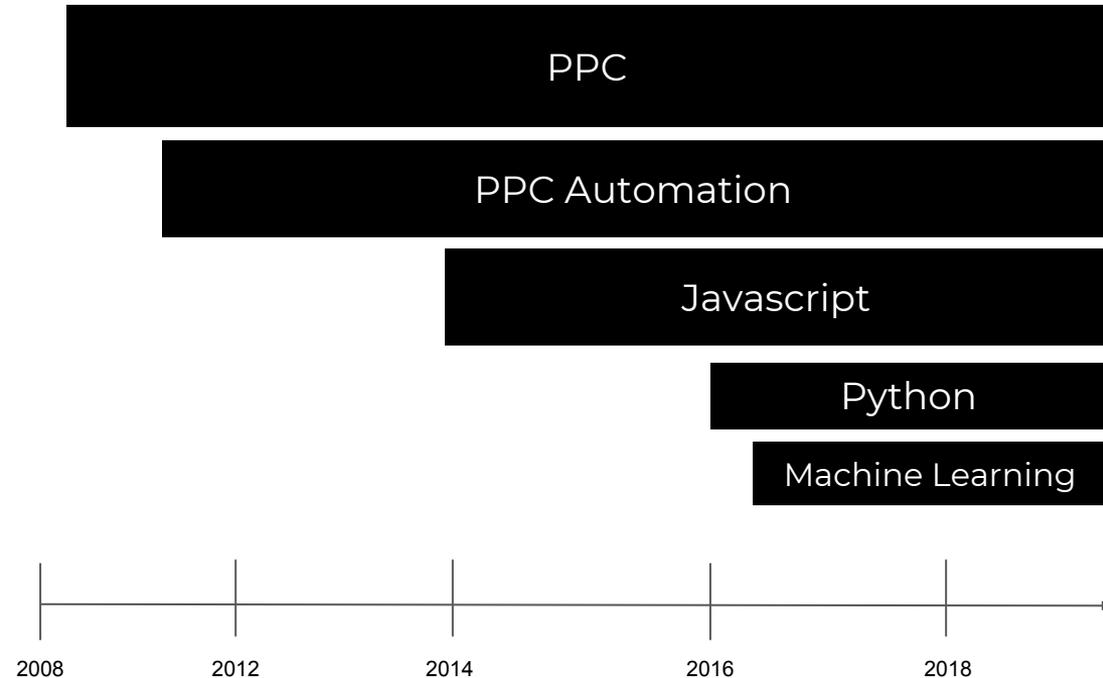


**Munich-based**

Focus Ecom & Retail

Self Taught Dev

Dad of 2,5 yr old



# Think of Me As a Knowledgeable Tourist



# Let's Get Started with ML Essentials

**INTRO**

Machine Learning Essentials



#FOS19

What Do We All Have in Common?

A scene from the movie Toy Story showing Woody and Buzz Lightyear. Woody is on the left, looking slightly concerned. Buzz is on the right, wearing his green and purple space suit and holding a purple ring on his finger. The background is a simple room with a door and a window.

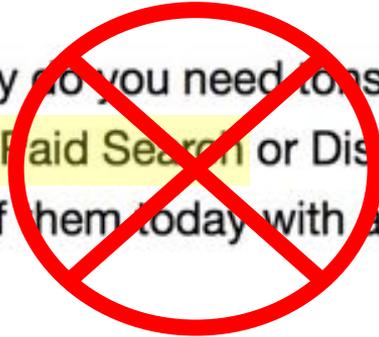
**DATA**

**DATA EVERYWHERE**

# Data-Driven Nature Makes ML Relevant



Why do you need tons of people at your agency/company to manage your **Paid Search** or Display Campaigns, when you can replace almost all of them today with a system that leverages ML?



# “ML-Worthy” PPC Automation Tasks



## Query Understanding

- Typo Detection
- Entity recognition
- ...



## Monitoring

- Anomaly detection
- Semantic inventory match
- ...



## Text Generation

- Keyphrase extraction
- Text summarization
- ...



Start with the problem, **not** the solution. Make sure you aren't treating ML as a hammer for your problems.

# Let's Get To Know our ML Starter Toolkit

INTRO

Machine Learning Essentials

**ML Toolkit For PPC**

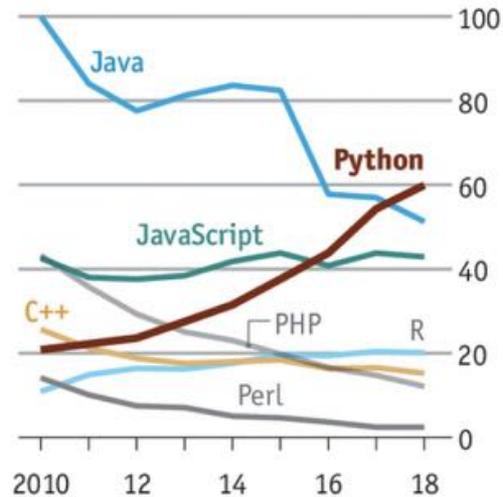
Platforms & Tools



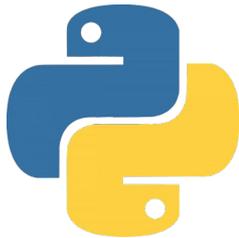
#FOS19

# Python: Most Popular in 2018 and Dutch!

US, Google searches for coding languages



Guido van Rossum

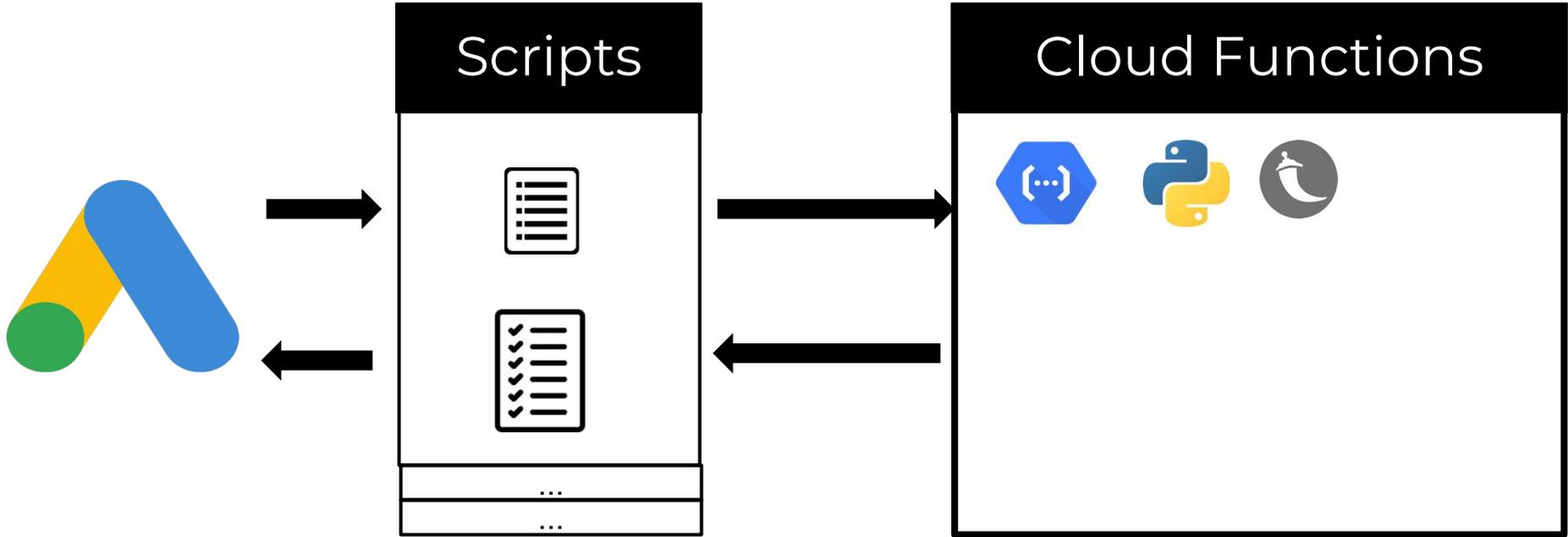


# Ads Scripts vs Python vs R?



	Ads Scripts	Python	R
Serverless	✓	(✓)	
Direct Ads API	✓		
Many Packages		✓	✓
General Purpose		✓	
Machine Learning		✓	✓

# No Either-Or: Run Python From Scripts



Demo: [bit.ly/norisk\\_python](https://bit.ly/norisk_python)

# Colaboratory: Google-Sheets for Python

```
CO norisk_FriendsOfSearch_PythonScripts.ipynb ☆
File Edit View Insert Runtime Tools Help
CODE TEXT CELL CELL

> 1.4 Knowledge Graph Search API (Google)
Documentation: https://developers.google.com/knowledge-graph/

1 import requests
2 import json
3
4 query = "cdg" # Haarlem, Grachten
5 language = "nl"
6 apiKey = "AIzaSyRkQJPoUYb25K6JwgQtDgIUGUmUowR7fgk"
7
8 kgUrl = "https://kgsearch.googleapis.com/v1/entities:search?query=" + query + "&languages=" + language
9 json_response = json.loads(requests.get(kgUrl).content)
10
11 print('All place results and their scores:\n')
12 [print(str(list['result'])) for list in json_response['itemListElement'] if 'Place' in list['result']]['0']
13
14 print('\n\n*****\n\nEntire JSON response:\n')
15 print('\n' + json.dumps(json_response, indent=2, sort_keys=False))
16
17 # Or view directly in Browser:
18 # https://kgsearch.googleapis.com/v1/entities:search?query=haarlem&languages=de&limit=10&indent=true&key
19

All place results and their scores:

{'#id': 'kg:/m/0qk79', '#type': ['Place', 'Thing', 'Airport', 'BusStation'], 'description': 'Luchthaven Schiphol', 'score': 0.9999999999999999}
{'#id': 'kg:/m/026svdt', 'name': 'Station Aéroport Charles-de-Gaulle 2 TGV', '#type': ['Place', 'Thing', 'Airport', 'BusStation'], 'description': 'Gare d\'aéroport Charles-de-Gaulle 2 TGV', 'score': 0.9999999999999999}
{'#id': 'kg:/m/0bk4cb', '#type': ['Thing', 'Place']}

*****

Entire JSON response:

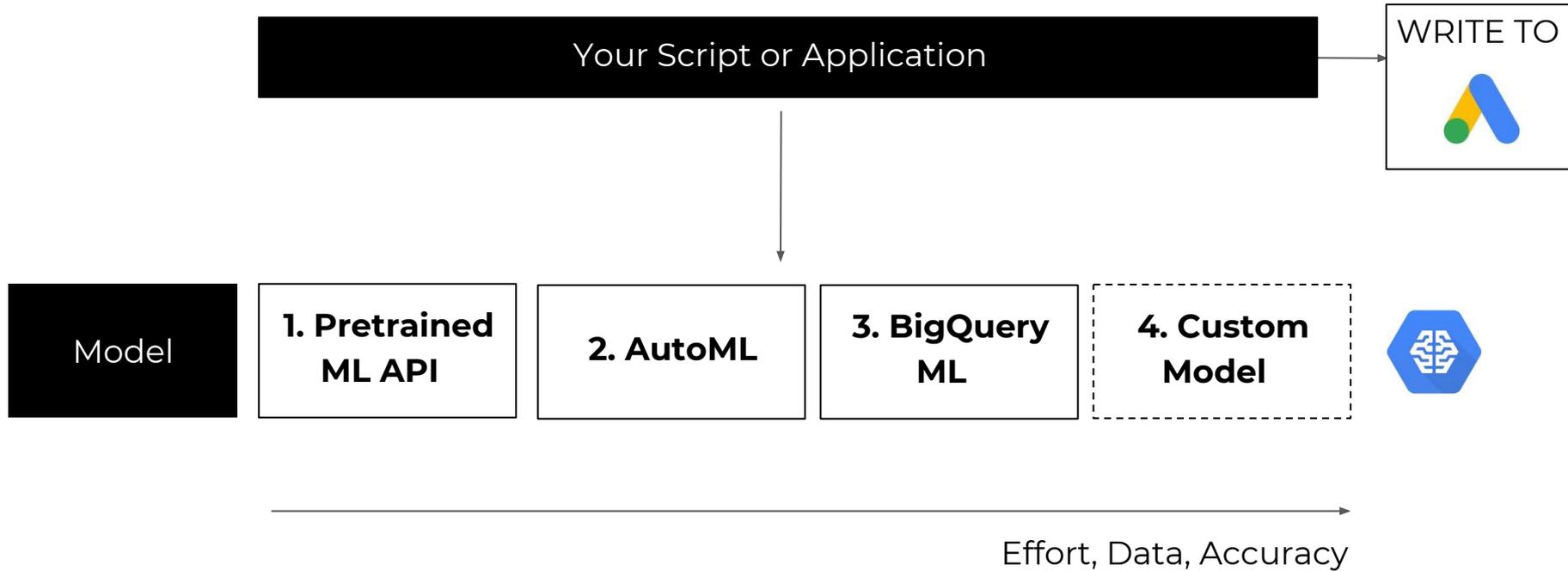
{
  "@context": {
    "vocab": "http://schema.org/",
    "goog": "http://schema.googleapis.com/",
    "EntitySearchResult": "goog:EntitySearchResult",
    "detailedDescription": "goog:detailedDescription",
    "resultScore": "goog:resultScore",
    "kg": "https://g.co/kg"
  },
  "itemListElement": [
    {
      "Place": {
        "#id": "kg:/m/0qk79",
        "#type": ["Place", "Thing", "Airport", "BusStation"],
        "description": "Luchthaven Schiphol",
        "score": 0.9999999999999999
      }
    },
    {
      "Place": {
        "#id": "kg:/m/026svdt",
        "name": "Station Aéroport Charles-de-Gaulle 2 TGV",
        "#type": ["Place", "Thing", "Airport", "BusStation"],
        "description": "Gare d'aéroport Charles-de-Gaulle 2 TGV",
        "score": 0.9999999999999999
      }
    },
    {
      "Thing": {
        "#id": "kg:/m/0bk4cb",
        "#type": ["Thing", "Place"]
      }
    }
  ]
}
```



**DEMO:**

[bit.ly/norisk\\_python](https://bit.ly/norisk_python)

# Three Simpler Options To Tie in ML



# Use Case #1: Classifying Near-Exact

INTRO

Machine Learning Essentials

ML Toolkit For PPC

Platforms & Tools

**Query Understanding**

1. Classifying Near-Exact



# Case #1: Classifying Near-Exact



Match the intent of a search with close variants



Search Engine Land

**When exact match isn't exact anymore: A script to regain control**

Contributor Daniel Gilbert



ONLY BLOCKING.

> What about **EXPANSION**?

# Classify Near Exact? Ask Google!

1. Suggest



```
← → ↻ suggestqueries.google.com/complete/search?output=toolbar&hl=en&q=topshpo+amsterdam
```

```
<toplevel/>
```

2. Custom Search



```
← → ↻ https://www.googleapis.com/customsearch/v1?gl=de&q=topshpo
```

```
{  
  kind: "customsearch#search",  
  - spelling: {  
    correctedQuery: "topshop",
```

Demo: [bit.ly/norisk\\_python](https://bit.ly/norisk_python)

# Auto ML I: Model scores & Prediction

Model  
campaignidentifi\_v20190124225001\_1

Created  
Jan 25, 2019  
3:15 AM

Analyzed  
44575 text items  
31 labels, 4547 test text items

Avg precision ⓘ  
0.888

Precision ⓘ  
81.695%

Recall ⓘ  
81.148%

plus size swim shorts

9979 characters remaining

PREDICT

Predictions

UP_US_SEA_PRO_Swimwear	0.955
UP_US_SEA_PRO_Bademode	0.426
UP_US_SEA_PRO_RLSA	0.226
UP_US_SEA_GEN_Mode	0.094
SU_US_SEA_GEN_Mode	0.017
UP_US_SEA_PRO_Bottoms	0.003
UP_US_SEA_BRA	0.002
UP_US_SEA_PRO_DUB_Struempfe	0.001

# Auto ML II: Classifying Brand Campaigns

## Confusion matrix

This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in orange).

True label	Predicted label	brand_Woolrich	brand_Dsquared2	brand_Windsor	brand_Herno	brand_Allude	brand_Caliban	G_Tracht_Designer	brand_Riani	brand_Moncler	brand_Parajumpers
brand_Woolrich	98.3%	1.2%	0.2%	-	-	-	-	-	0.3%	-	-
brand_Dsquared2	0.5%	99.2%	-	-	-	-	-	-	0.3%	-	-
brand_Windsor	0.7%	2.0%	96.6%	-	-	-	-	0.7%	-	-	-
brand_Herno	-	3.1%	-	94.4%	0.6%	-	-	-	1.9%	-	-
brand_Allude	-	1.9%	-	-	97.6%	-	-	0.5%	-	-	-
brand_Caliban	-	17.4%	-	-	-	82.6%	-	-	-	-	-
G_Tracht_Designer	0.7%	2.1%	-	-	-	-	94.4%	2.8%	-	-	-
brand_Riani	0.2%	0.2%	-	-	-	-	-	99.6%	-	-	-
brand_Moncler	0.7%	0.5%	-	-	-	-	-	0.2%	98.6%	-	-
brand_Parajumpers	1.3%	0.4%	-	-	-	-	-	-	-	98.2%	-

# Use Case #2: Ngram Analysis & BigQuery

INTRO

Machine Learning Essentials

ML Toolkit For PPC

Platforms & Tools

Query Understanding

1. Classifying Near-Exact

**Prediction**

2. Analysing Query n-Grams



# Remember This? BigQuery Is Faster!

Search Engine Land

## Ads Script: Find Your Best And Worst Search Queries Using N-Grams

Former Googler Daniel Gilbert of Brainlabs shares a script to help you identify wasted Ads spend using n-grams.

Daniel Gilbert on August 26, 2015



# The Basis: Google Ads Data Transfer

The screenshot shows the Google BigQuery interface for configuring a new data transfer. On the left, there is a sidebar with navigation options: 'COMPOSE QUERY' (highlighted in red), 'Query History', 'Job History', 'Scheduled Queries', and 'Transfers'. The main area is titled 'New Transfer' and includes a link to learn more about configuring Google Ads (formerly AdWords). The configuration fields are as follows:

- Source:** Google Ads (formerly AdWords)
- Display name:** NL
- Refresh window:** 30
- Schedule:** every 24 hours (with 'Reset' and 'Edit' buttons)
- Destination dataset:** crawler\_test
- Customer ID:** 123-4567-890
- Exclude removed/disabled items:**

At the bottom, there is an 'Advanced' section and two buttons: 'Add' (in blue) and 'Cancel'.

The screenshot shows a list of tables under the 'SPOR\_Transfer' dataset. Each table name is preceded by a grid icon and a three-dot menu icon. The tables listed are:

- AccountBasicStats\_2131444427
- AccountConversionStats\_21314444...
- AccountNonClickStats\_2131444427
- AccountStats\_2131444427
- Ad\_2131444427
- AdBasicStats\_2131444427
- AdConversionStats\_2131444427
- AdCrossDeviceConversionStats\_21...
- AdCrossDeviceStats\_2131444427
- AdGroup\_2131444427
- AdGroupBasicStats\_2131444427
- AdGroupConversionStats\_2131444...
- AdGroupCrossDeviceConversionSta...
- AdGroupCrossDeviceStats\_213144...
- AdGroupStats\_2131444427
- AdStats\_2131444427
- AgeRange\_2131444427

# BigQueryML: Train Models with SQL

Demo: [bit.ly/norisk\\_python](https://bit.ly/norisk_python)

# Our Agenda: Intro & 3 PPC Use Cases

INTRO

Machine Learning Essentials

ML Toolkit For PPC

Platforms & Tools

Query Understanding

1. Classifying Near-Exact

Prediction

2. Analysing Query n-Grams

**Text Summarization**

3. Finding Key Phrases

# Text Summarization > see Demo

## ▾ **PART 3. TEXT SUMMARIZATION**

### ▸ **3.1 Text Summarization with Sumy (Dan Shapiro)**

↳ 2 cells hidden

### ▸ **3.2 KeyPhrase Detection API (Microsoft Azure)**

↳ 2 cells hidden

Demo: [bit.ly/norisk\\_python](https://bit.ly/norisk_python)

# Takeaways: Start Experimenting!

**INTRO**

Machine Learning Essentials

**ML Toolkit For PPC**

Platforms & Tools

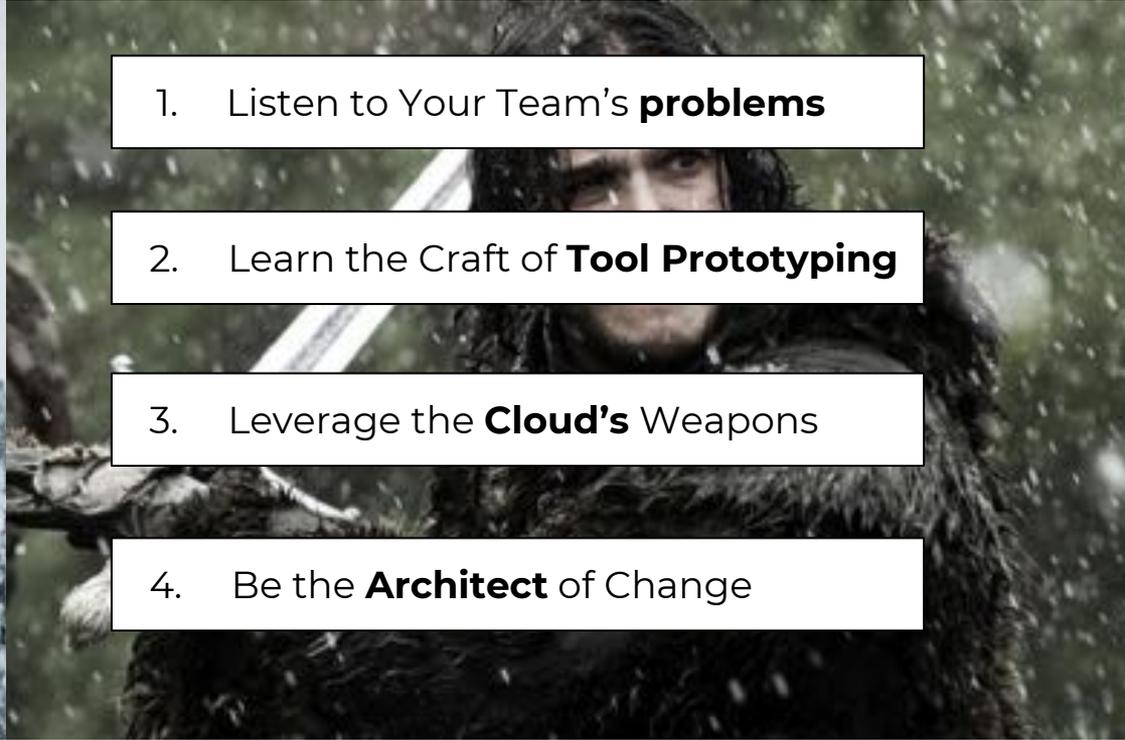
**USE CASES**

Typos, NGrams, Summaries

**TAKEAWAYS**

Start Experimenting!

# AI Is Coming ? Let's be Jon Snow!



1. Listen to Your Team's **problems**

2. Learn the Craft of **Tool Prototyping**

3. Leverage the **Cloud's** Weapons

4. Be the **Architect** of Change

# THANK YOU. Your Questions Please!

## #FOS19

### MACHINE LEARNING IN PPC

How To Get Started Today

Christopher Gutknecht | norisk Group | #FOS19



# Bonus: GA Anomaly Detection for Slack

#om\_ga\_alarmduck

☆ | 6 | 0 | Thema hinzufügen

Freitag, 4. Januar



Alarmduck APP 07:23 Uhr

Anomalies in |

Page: /

Bounce Rate: 10.5879650896%

▲ 56% UP compared to expected value (46 kB) +



07:23 Uhr Anomalies in |

Page: /Wintersport/Skitouren/

Pageviews: 321

▲ 78% UP compared to expected value (31 kB) +



# Our Agenda: Intro & 3 PPC Use Cases

**INTRO**

Machine Learning Essentials

**ML Toolkit For PPC**

Platforms & Tools

**USE CASES**

Typos, NGrams, Summaries

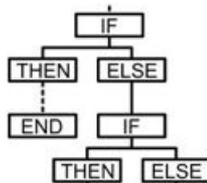
**APPENDIX I**

Additional Content

# Rule #1: Not Every Problem Needs ML

## RULE BASED

- Designed rule-flow
- **Outliers not included**
- Won't improve on data



## MACHINE LEARNING

- Algorithmic model
- **Model learns from errors**
- Model constantly retrained

$$E = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

# Rule-Based vs Machine Learning Code

Write a computer program  
with **explicit rules** to follow

```
if email contains V!agrà  
    then mark is-spam;  
if email contains ...  
if email contains ...
```

**Traditional Programming**

Write a computer program  
to **learn from examples**

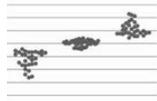
```
try to classify some emails;  
change self to reduce errors;  
repeat;
```

**Machine Learning Programs**

# Five Main Types of ML Algorithms

Unsupervised

Clustering



Dimension Reduction



Supervised

Classification



Regression

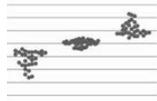


+ Reinforcement Learning

# Natural Language Processing Is Its Own Game

Supervised ← → Unsupervised

Sentiment Analysis



Topic Modelling



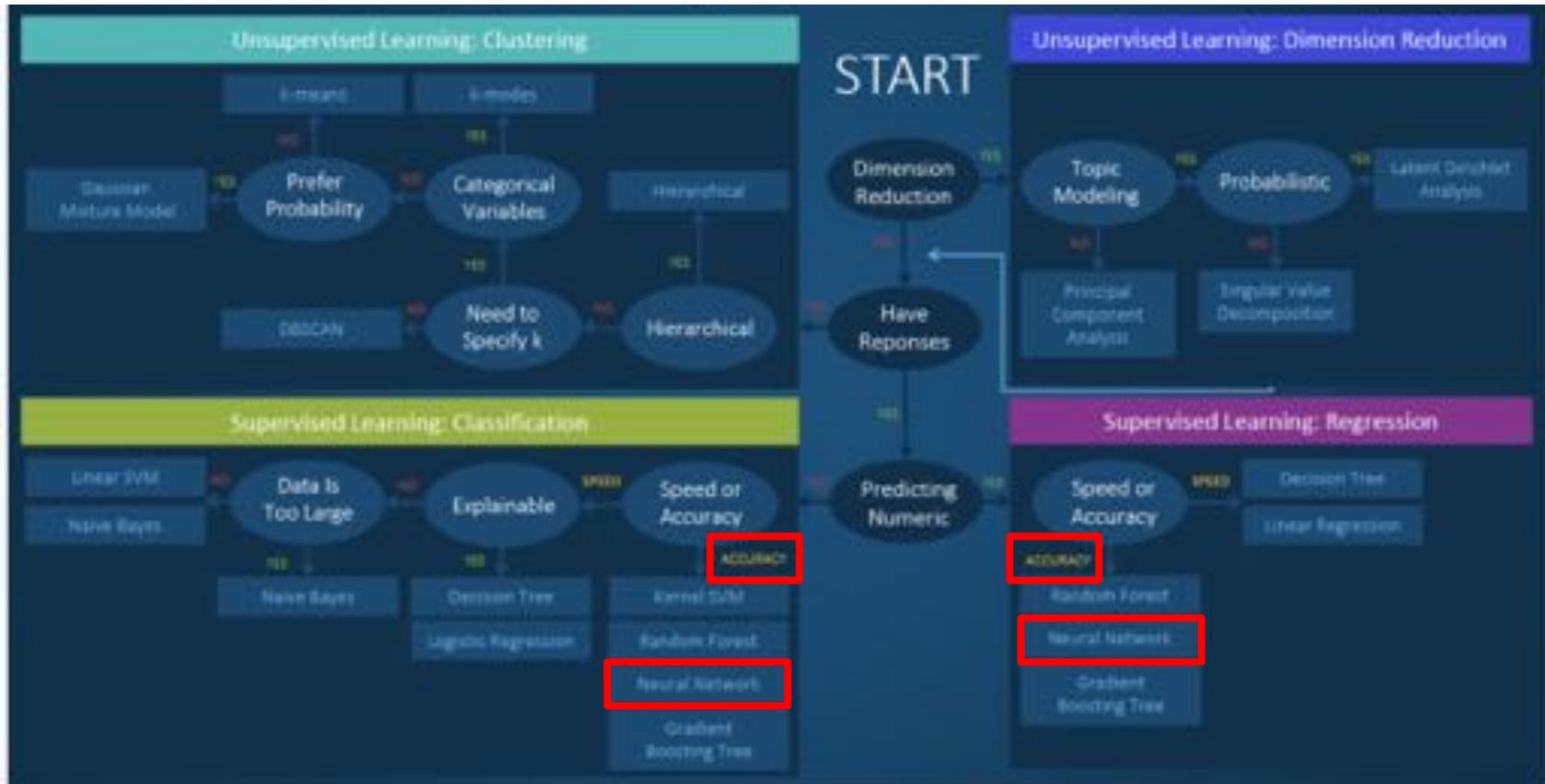
Entity Recognition



Sequence Prediction



# And Deep Learning? Higher Accuracy!



# Main Challenges for ML Approaches

## Good Problem Framing

- Finding value drivers
- Defining relevant outputs



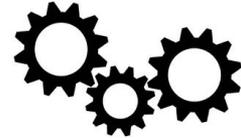
## Access To Data

- Depth and Breadth
- Cleanliness



## Knowledge of Solutions

- Choice of framework
- References projects



# The 3 Cloud Platforms & their ML Tools



## Google Cloud Platform

+ Google Integrations

- Managed **Storage** 
- Serverless **Execution** 
- Pretrained ML **APIs** 
- **AutoML** Service 

## Amazon Web Services

+ Best DevOps Workflow

- “
- “
- “
- “

## Microsoft Azure

+ Windows & Bing maybe

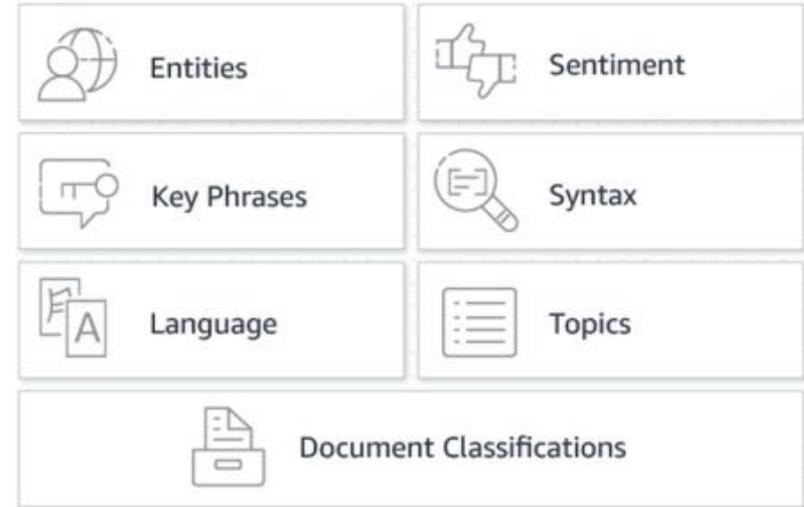
- “
- “
- “
- -

→ All similar. But Google has its benefits!

# Pretrained APIs: Amazon Comprehend



The screenshot shows the AWS console interface for Amazon Comprehend. At the top, there is the AWS logo, followed by 'Services' and 'Resource Groups' dropdown menus. Below this, the text 'Machine learning' is displayed. The main heading is 'Amazon Comprehend' in a large, bold font, followed by 'Natural Language Processing and Text Analytics'. A descriptive paragraph states: 'Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text.'



Extracts data, topics, and document classifications with confidence scores

# Look out for Colab: Example TensorFlow

TensorFlow

Learn

TUTORIALS GUIDE DEPLOY

Get started with TensorFlow

Learn and use ML

- Overview
- Basic classification
- Text classification
- Regression
- Overfitting and underfitting
- Save and restore models

## Text classification with movie reviews

 [Run in Google Colab](#)

This notebook classifies movie reviews as *positive* or *negative* using the text of or two-class—classification, an important and widely applicable kind of machine

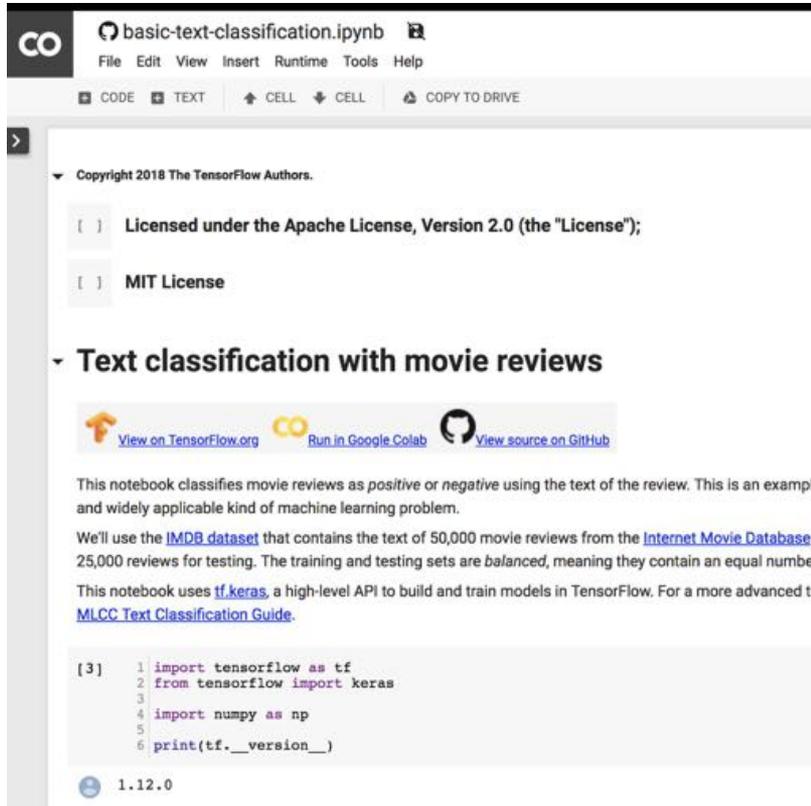
We'll use the [IMDB dataset](#) that contains the text of 50,000 movie reviews from split into 25,000 reviews for training and 25,000 reviews for testing. The training they contain an equal number of positive and negative reviews.

This notebook uses [tf.keras](#), a high-level API to build and train models in TensorFlow classification tutorial using `tf.keras`, see the [MLCC Text Classification Guide](#)

```
import tensorflow as tf
from tensorflow import keras

import numpy as np

print(tf.__version__)
```



The screenshot shows the Google Colab interface for a notebook titled 'basic-text-classification.ipynb'. The top navigation bar includes 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. Below this are buttons for 'CODE', 'TEXT', 'CELL', and 'COPY TO DRIVE'. The notebook content is displayed in a dark-themed editor. It starts with a copyright notice for TensorFlow Authors, followed by license information: 'Licensed under the Apache License, Version 2.0 (the "License");' and 'MIT License'. A section header 'Text classification with movie reviews' is followed by a toolbar with icons for 'View on TensorFlow.org', 'Run in Google Colab', and 'View source on GitHub'. The main text of the notebook describes the task of classifying movie reviews as positive or negative using the IMDB dataset. It mentions the use of tf.keras and provides a link to the MLCC Text Classification Guide. A code cell is shown with the following Python code:

```
[3] 1 import tensorflow as tf
    2 from tensorflow import keras
    3
    4 import numpy as np
    5
    6 print(tf.__version__)
```

Below the code cell, the output shows the TensorFlow version: 1.12.0.

# Popular Python NLP Packages

**fuzzyWuzzy**

**spaCy**

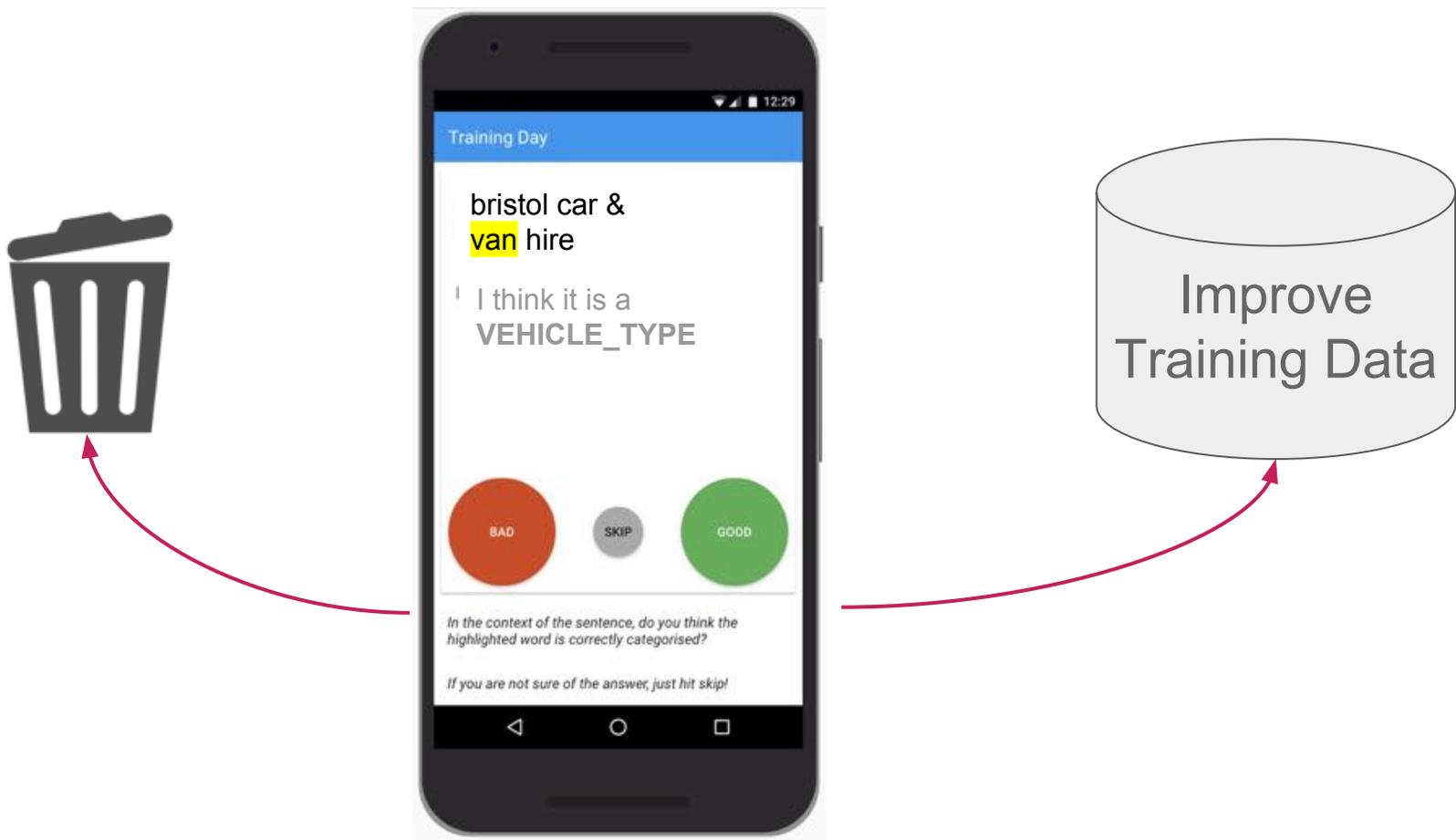
**gensim**

String Similarity $\approx$	Language Model 	Context Modeling $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
		
		
		

Fuzzy Wuzzy Cloud Function Example:

<https://colab.research.google.com/drive/1W6guxVZJqp-duQT6jSlmsqxkrCgt95Cl#scrollTo=a60zDUMiOEXD>

# Sixt: Query Disambiguation via App & Spacy

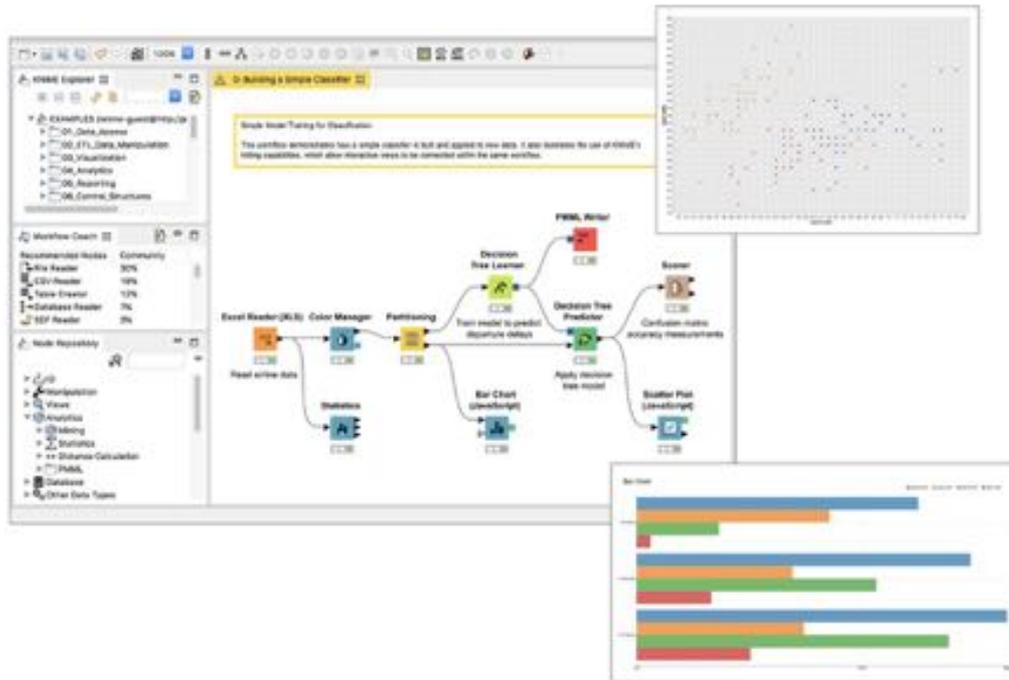


# Don't want write code? Try KNIME



## KNIME Analytics Platform

Open, intuitive, integrative data science.



## Querying Google Analytics in KNIME

Mon, 10/06/2014 - 00:00 — winter

The KNIME Google API extension (since [version 2.10](#)) allows for the connection and interaction of KNIME with Google APIs. For now nodes are provided to request and load data from [Google Analytics](#).



# Our Agenda: Intro & 3 PPC Use Cases

**INTRO**

Machine Learning Essentials

**ML Toolkit For PPC**

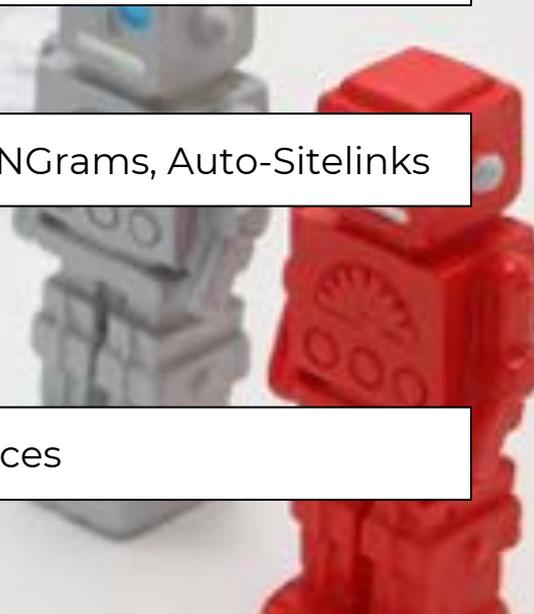
Platforms & Tools

**USE CASES**

Typos, NGrams, Auto-Sitelinks

**APPENDIX II**

Resources



# Resources: Google ML Crash Course



[developers.google.com/machine-learning/crash-course/](https://developers.google.com/machine-learning/crash-course/)

# Resources: [ai.google/education](https://ai.google/education)

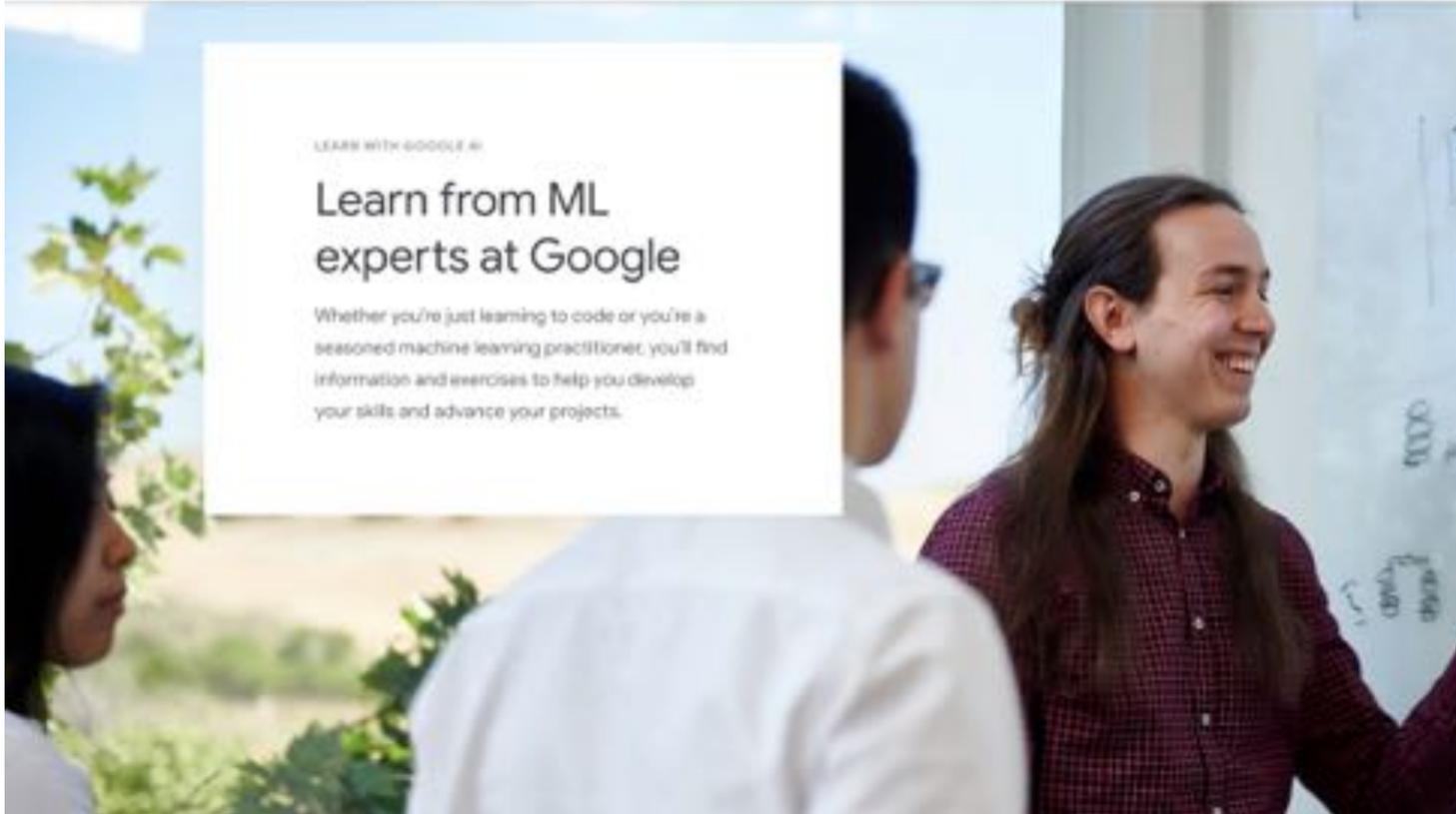
Google AI

About Stories Research Education Tools Principles Blog

LEARN WITH GOOGLE AI

## Learn from ML experts at Google

Whether you're just learning to code or you're a seasoned machine learning practitioner, you'll find information and exercises to help you develop your skills and advance your projects.



# Resources: cloud.google.training

» Training home

## Training Programs

Learning Tracks

Cloud Infrastructure

Data & Machine Learning

Application Development

G Suite Administration

Your Gateway to GCP

Course Catalog

Find a Class

Self-Paced Training

Qwiklabs

Courses

## Data and Machine Learning

This learning path is designed for data professionals who are responsible for designing, building, analyzing, and optimizing big data solutions. To get up to speed quickly, choose a course track suited for your role or interests.

### Data Analyst Track

STEP  
1

#### From Data to Insights with Google Cloud Platform

This course teaches participants how to derive insights through data analysis and visualization using the Google Cloud Platform.

[ON DEMAND](#) [MORE INFO](#)

STEP  
2

#### Building Conversational Experiences with Dialogflow

This course provides a deep dive into how to create a chatbot using Dialogflow, augment it with Cloud Natural Language API, and operationalize it using Google Cloud tools.

[ON DEMAND](#) [MORE INFO](#)

# Resources: [codelabs.google.com](https://codelabs.google.com)

The screenshot shows the Google Developers Codelabs page. At the top, there is a search bar and the text "Google Developers". Below this, there is a grid of 12 lab cards, each with a title, duration, update date, a "Start" button, and the Google Cloud logo.

Lab Title	Duration	Updated
Building a Serverless Data Pipeline: IoT to Analytics	44 min	Updated Nov 8, 2019
Building a gRPC service with C#	31 min	Updated Nov 20, 2019
Building a gRPC service with Java	22 min	Updated Nov 14, 2019
Building a gRPC service with Node.js	44 min	Updated Jun 26, 2017
Classify Text into Categories with the Natural Language API	33 min	Updated Nov 25, 2019
Classify images of clouds in the cloud with AutoML Vision	55 min	Updated Nov 24, 2019
Compute the Cosmos with Google Compute Engine	54 min	Updated Nov 11, 2019
Configure an Uptime Check and Alerting Policy	21 min	Updated Nov 13, 2019
Connect and visualize all your data in Data Studio	31 min	Updated Nov 5, 2019
Connecting to Cloud SQL	34 min	Updated Dec 26, 2019
Continuous Delivery to Kubernetes Using Spinnaker	51 min	Updated Oct 25, 2019
Continuous Deployment with Cloud Build	34 min	Updated Nov 5, 2019



# Resources: aws.training

The screenshot displays the AWS Training and Certification portal. The top navigation bar includes 'Dashboard', 'Learning Library', 'Certification', and 'Support'. The user interface is in English and shows 'My Account'. The main content area is titled 'View All' and is filtered to 'Digital Training'. A sidebar on the left lists various domains, with 'Machine Learning' selected. The main grid features nine resource cards:

- Amazon Transcribe Deep Dive: Using Feedback Loops to Improve** (Video, Advanced, 43 minutes)
- What is Deep Learning?** (Video, Fundamental, 20 minutes)
- What is Machine Learning?** (Video, Fundamental, 10 minutes)
- What is Artificial Intelligence?** (Video, Fundamental, 10 minutes)
- Machine Learning Security** (Curriculum, Intermediate)
- Machine Learning Exam Basics** (Curriculum, Certification)
- Conversation Primer: Machine Learning Terminology** (Curriculum, Fundamental)
- Developing Machine Learning Applications** (Curriculum, Advanced)
- ML Building Blocks: Services and Terminology** (Curriculum, Fundamental)

# Resources: Machine Learning 101



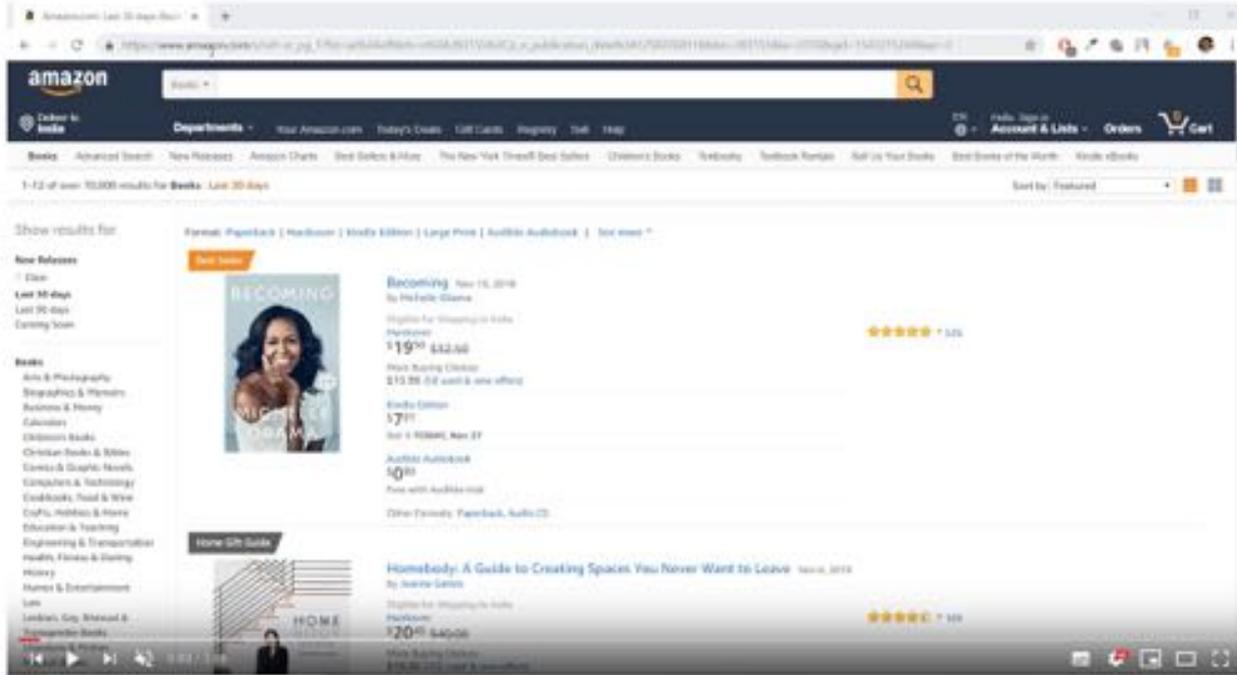
[docs.google.com/presentation/d/1kSuQyW5DTnkVaZEjGYCkfOxvzCqGEFzWBy4e9Uedd9k/](https://docs.google.com/presentation/d/1kSuQyW5DTnkVaZEjGYCkfOxvzCqGEFzWBy4e9Uedd9k/)

# Resources: TechSEO Boost (Python/ML)



Python & SEO Talks @ TechSEOBoost: [https://youtu.be/N0uJp\\_JXfOg](https://youtu.be/N0uJp_JXfOg)

# Resources: Scraping with Python



Python  
Python Scrapy Tutorial- 1 - Web Scraping Spiders and Crawling

Python Scrapy Tutorial: [https://youtu.be/ve\\_Oh4Y8nuI](https://youtu.be/ve_Oh4Y8nuI)