



Achieve More with Less using Active Learning & Guided Search

Data Science Summit 2018 Workshop.

Omri Allouche & Tal Perry

Dr. Omri Allouche

Head of Research, Gong.io

Teacher, Bar Ilan University





0,000,000



Record



Transcribe



Analyze

Gong.io // ACME - Demo Call

Customer

- ACCOUNT NAME: ACME
- OPPORTUNITY: ACME
- STAGE DURING CALL: Meeting
- STAGE NOW: Closed Lost
- AMOUNT: \$4,800
- > 2 CALLS WITH NOVETTA

Call Details

- HOST: Amit Bendov
- RECORDED: Nov 18, 2016 10:30 A...

Trackers

- QUANTITIES: 100 calls

REP SELF COACHING

- Improve their own pitch
- Follow up on action items
- Prepare for the next call

Search for Keywords

Search for anything said in the call

Save calls to libraries

Track key moments

Automatically identify topics

Monitor the Talk-to-Listen ratio

Comment on calls

Chris Orlob 11:54 | Great demonstration of transcription keyword search h

Chris Orlob 13:32 | Zackary Simkover check this question for your next call

Chris Orlob 19:16 | Explain this more deeply next time.

19:01

Enter comment. (Ctrl+Enter to post)

Post Comment

Others are notified. Go to loop them in.

63%

37%

Amir Bendov CEO, Gong

Linda Crump VP Sales

Artificial Intelligence Phone + Conferencing Recording Machine Learning Salesforce.com Search + Trackers Training Pricing Trial Setup

Teaching Machines is an Expensive Business...

- Transcribing an hour of Speech \approx \$100
- You need roughly 1,000 hours for descent performance = \$100,000
- Labeling a text sentence in FigureEight is \sim 10 cents
- Binary Classification Tasks with \sim 1-5% prevalence of target class \times 1,000 sentences = \$10,000 per task
- Annotating Sales calls requires domain expertise – hard to come by and expensive...



Action Items Detection

I will send you an email after the call

Let's touch base next Tuesday

Let me share my screen...



Other features can indicate an Action Item:

- Action items are mostly said by Sales Rep
- Action items mostly appear at the end of the call
- Action items are usually the last sentence in a paragraph

Action Items Detection

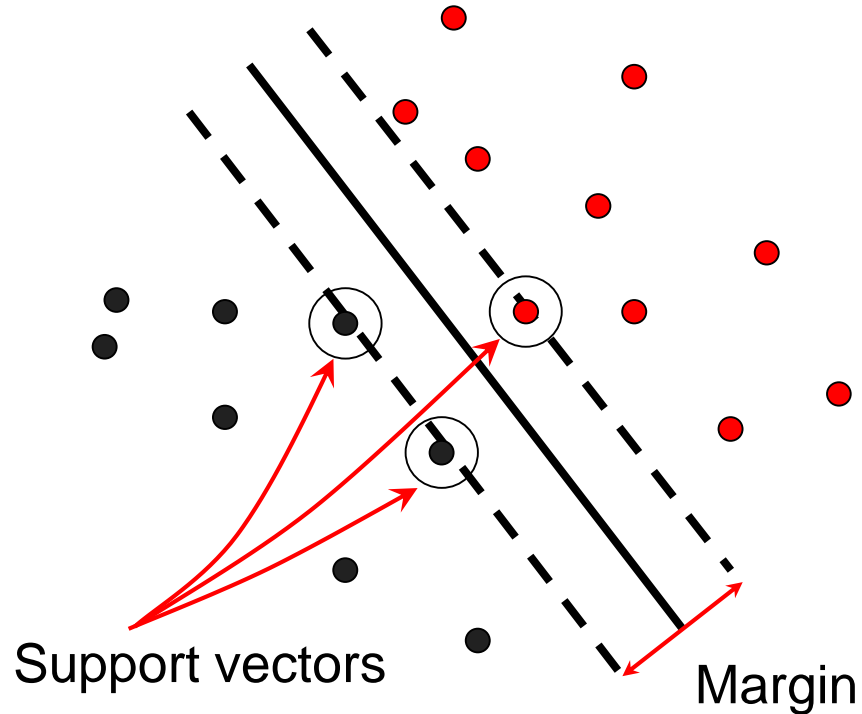
- Many sentences are obviously not an Action Item –
 - *Hey Josh. How are you?*
 - *Yeah, it's been raining all day here.*
 - *Can you hear me?*
 - ...
- We don't want to spend precious labeling time on them
- **Claim:** Even a basic model can detect if a sentence is:
 - Definitely an Action Item
 - Definitely not an Action Item
 - Might be an Action Item



Concepts routed in other Machine Learning concepts

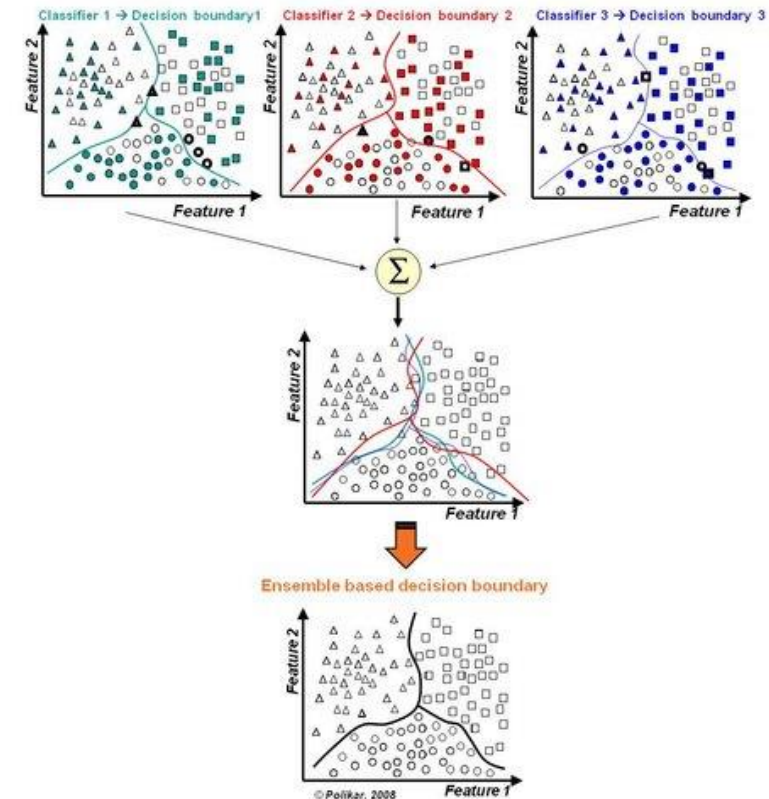
SVM - Support Vector Machines –

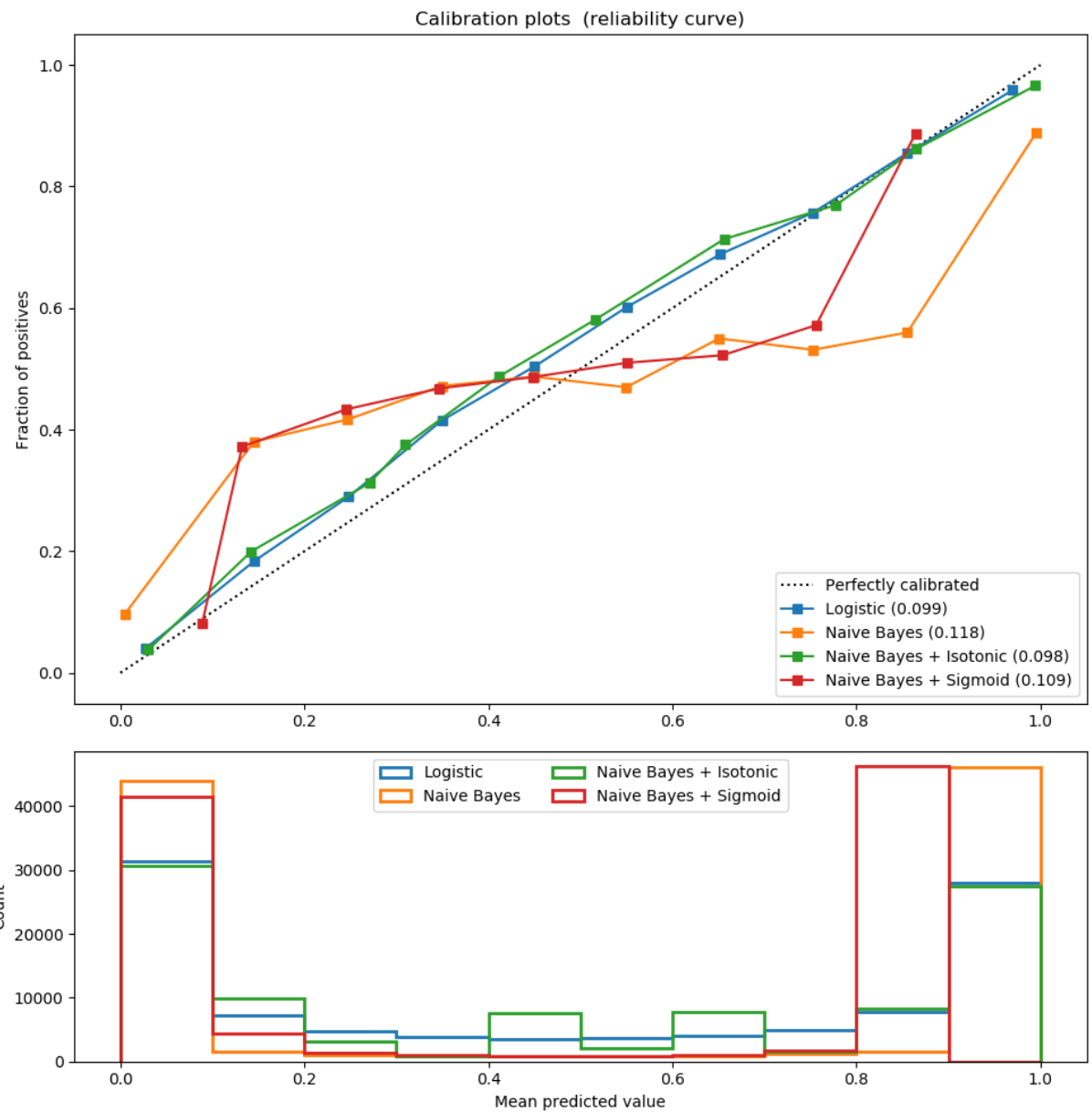
Decision boundary is set by Support Vectors –
these points are the most informative



Ensemble Learning –

Combine different models for a joint prediction, with
higher accuracy.

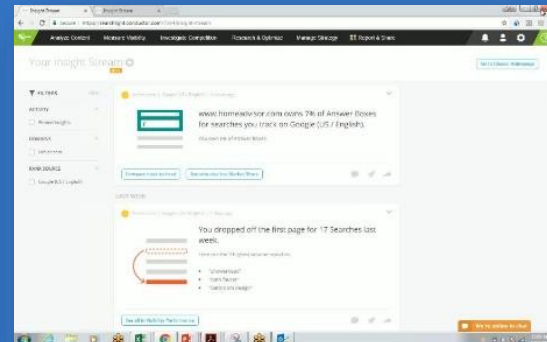




Video Frame Classification

Video recordings can be separated into 4 main categories:

1. Web cam
2. Browser demo
3. Presentation slides (ppt / Google slides etc)
4. Conferencing provider app (Skype / Zoom / GoToMeeting etc)



Pitfalls get more difficult as complexity grows



Impact

- Overly dependent on IT
- Can't get immediate answers
- High TCO for complex data

What makes us unique?

Choices in Market Data



Modern Surveys

- Cross-sectional data from over 34 million profiles
- 15,000+ jobs
- Real time data & historical trends based on actual data
- Because data is cross-sectional, actually benefits from the decline of survey
- Robust data for hundreds of skills, degrees, occupations, hot jobs, etc.



Data Aggregators

- Based on survey data - aged and adjusted by differentials, premiums, and discounts
- Dependent on underlying surveys, but requires surveying itself - measurements declining survey participation rates
- No ability to see actual data trends, only "aged" & "suppressed" trends
- No data on pay for skills, advanced degrees, hot jobs, etc.
- Fast to find "a number" for 4,000+ jobs



Traditional Surveys

- Long trusted source of market data
- At least 6 months old when published, often wait for a full year beyond that - 18 months old
- Usually need multiple sources to get to 80% coverage on occupations
- Participation rates and data volumes on the decline in the last 10 years
- No data on pay for skills, advanced degrees, hot jobs, etc.

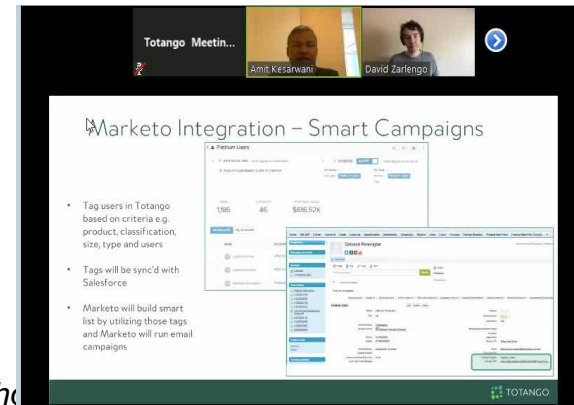
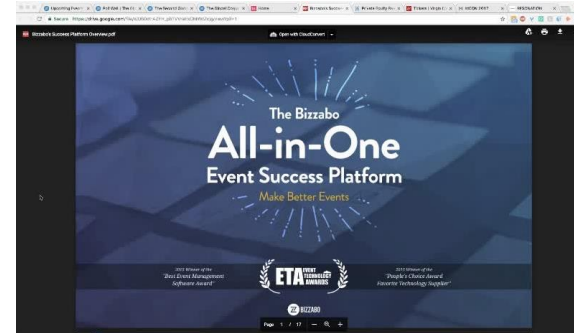
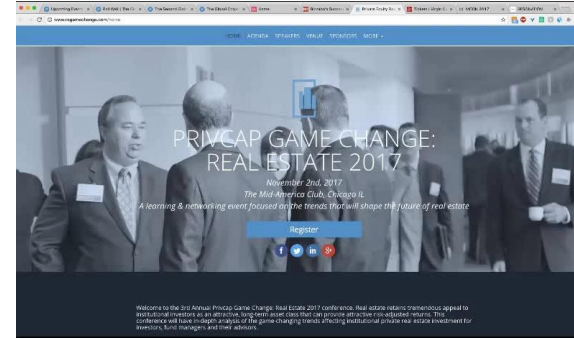
ANALYTICS & INSIGHTS DELIVERING MEASURABLE ROI

Delivers communication insights to improve the employee experience that HR has no way of measuring today.

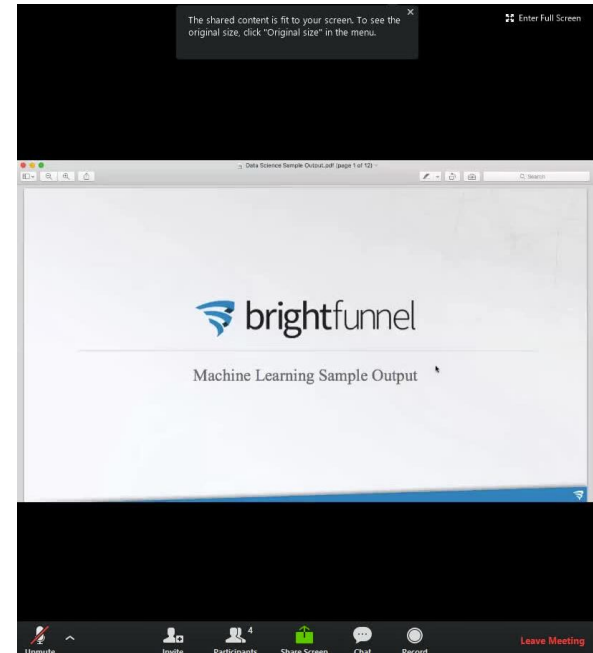
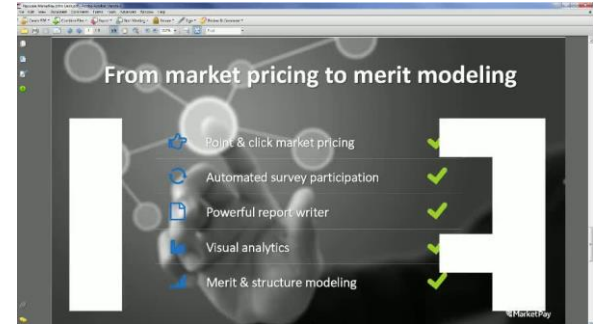
- Analytics on how, when and what information employees consume
- Aggregated data across all customers (billions of actions each year)
- Measures level of content engagement to drive future actions
- Test practices and insights learned from 4M+ employee interactions



Identifying Decision Makers A LONG AND COSTLY PROCESS...



Low Confidence



High Confidence



Stronger data, innovative compensation solutions

Mandy Lindgren, CCP
Account Executive

sh

CrowdFlower changes its name to Figure Eight

“

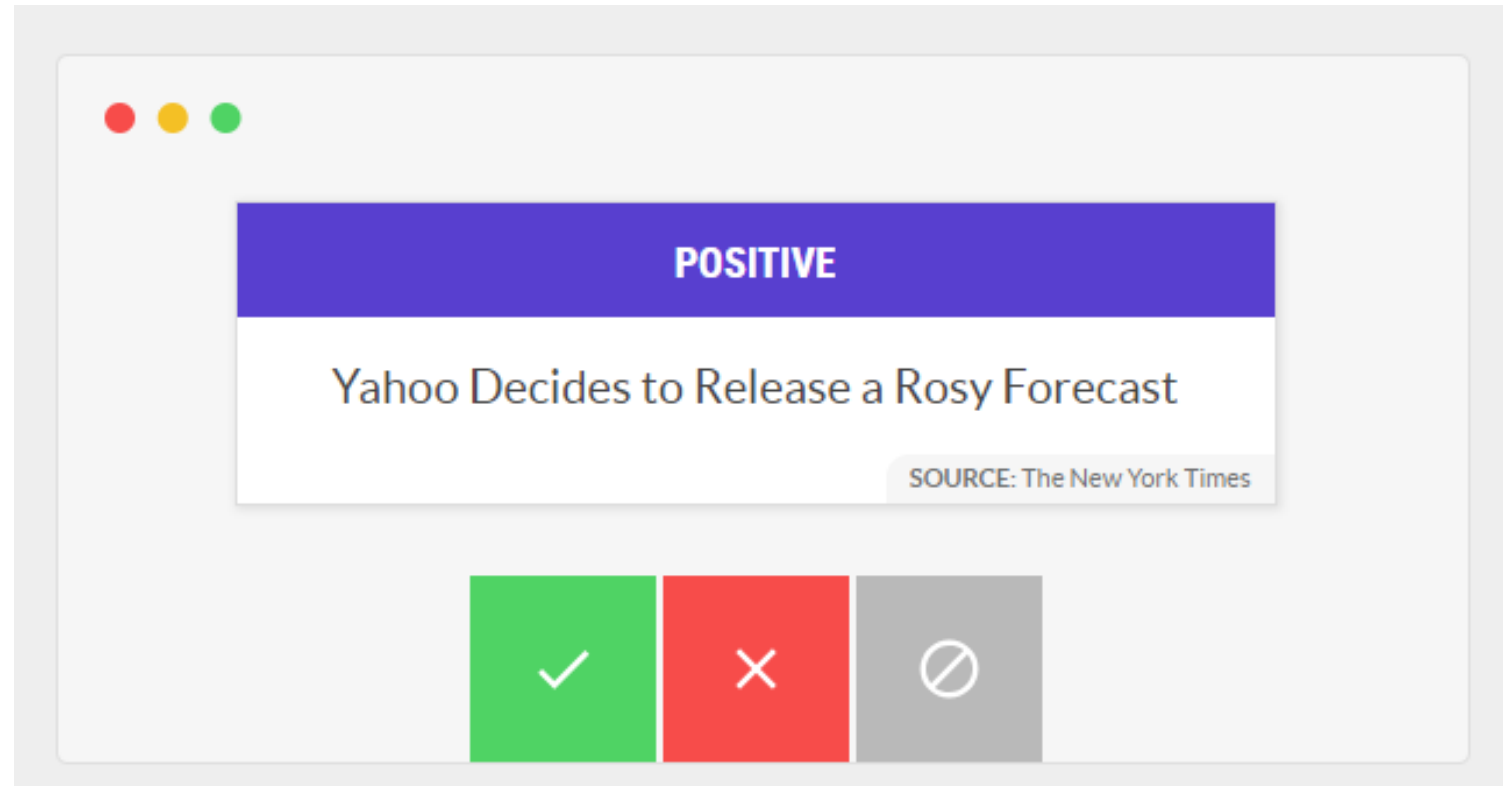
The formal name change to Figure Eight is to better reflect the company's technology platform that combines the best of machine learning and human intelligence, rather than just using human intelligence to label data. The new name and identity brings into focus Figure Eight's core philosophy of human-in-the-loop practices being the essential ingredient to making AI work in the real world, specifically the iterative process of active learning for the training, testing, and tuning of machine learning models.

”

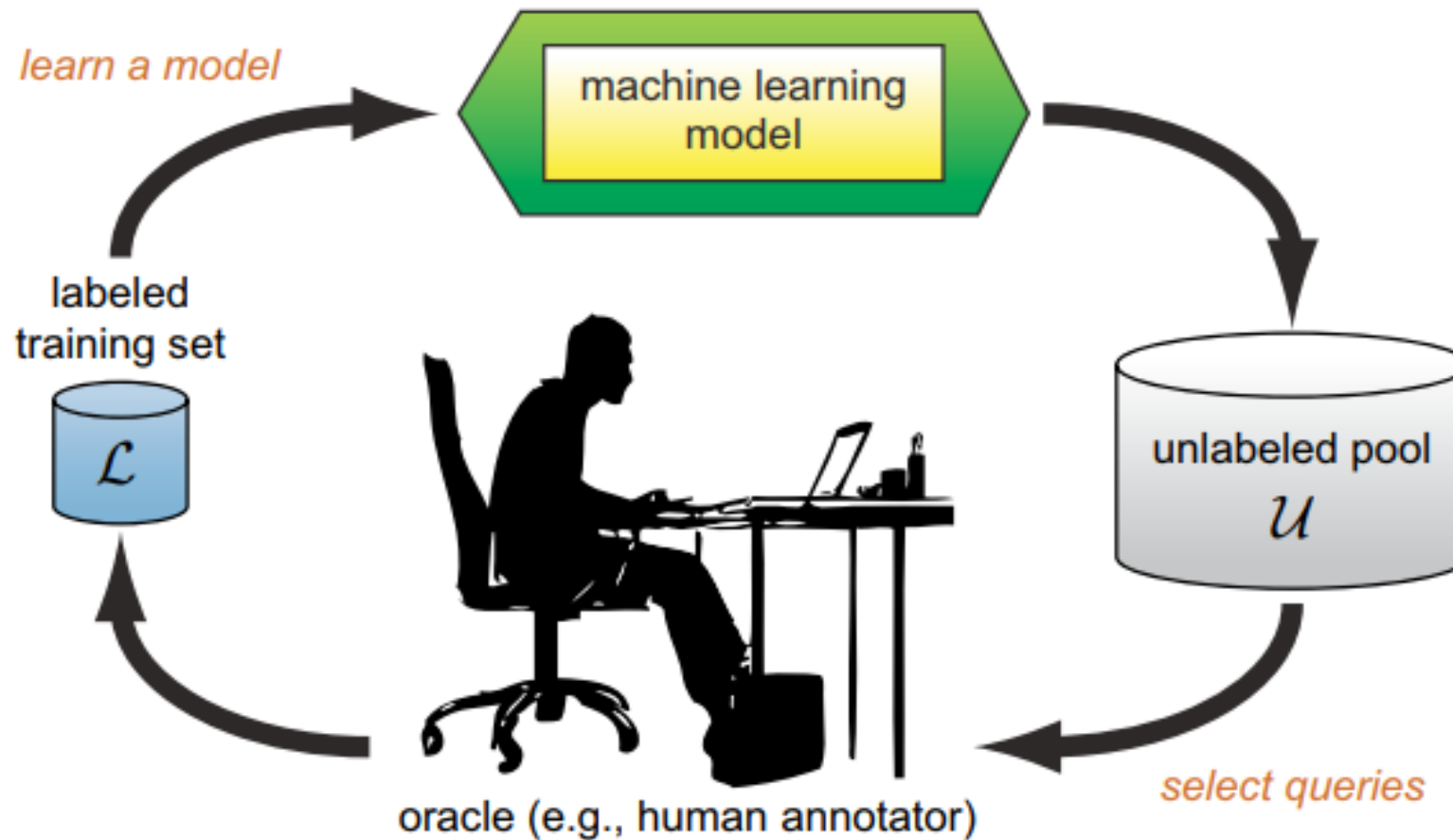
prodigy

Radically efficient machine teaching.
An annotation tool powered
by active learning.

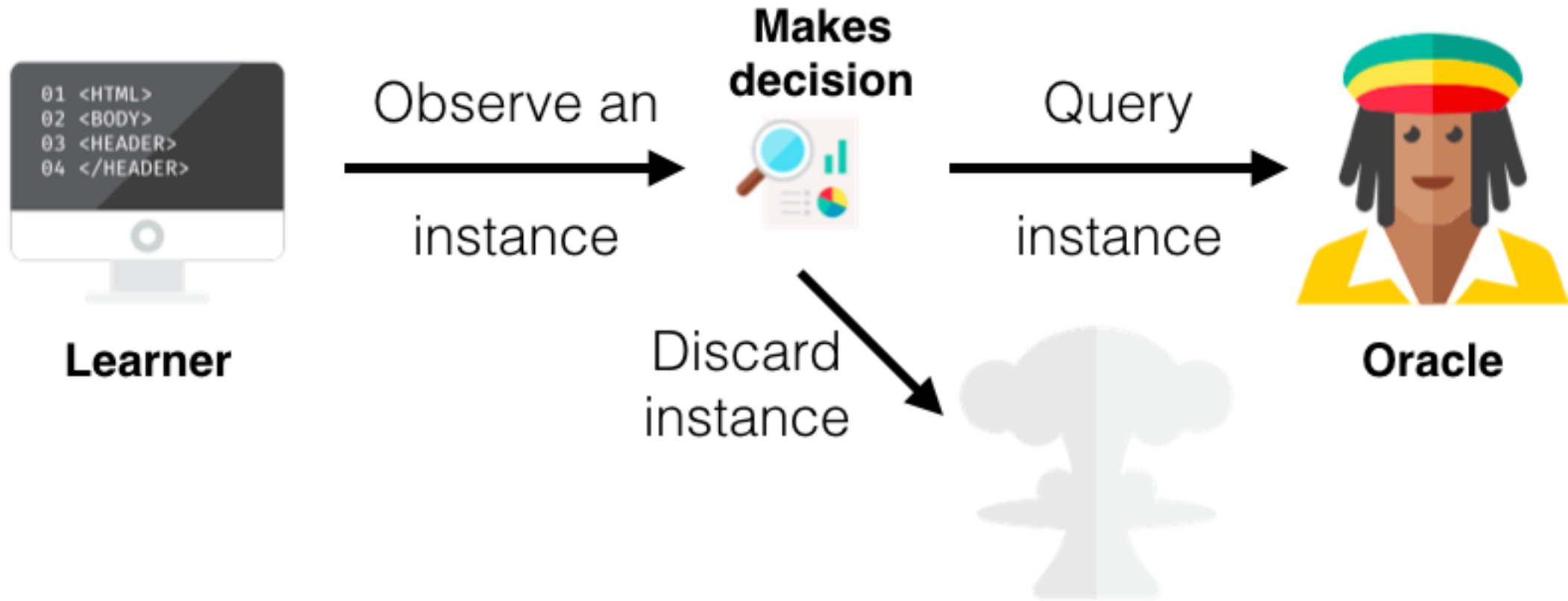
FROM THE MAKERS OF SPACY



Active Learning – Basic Framework



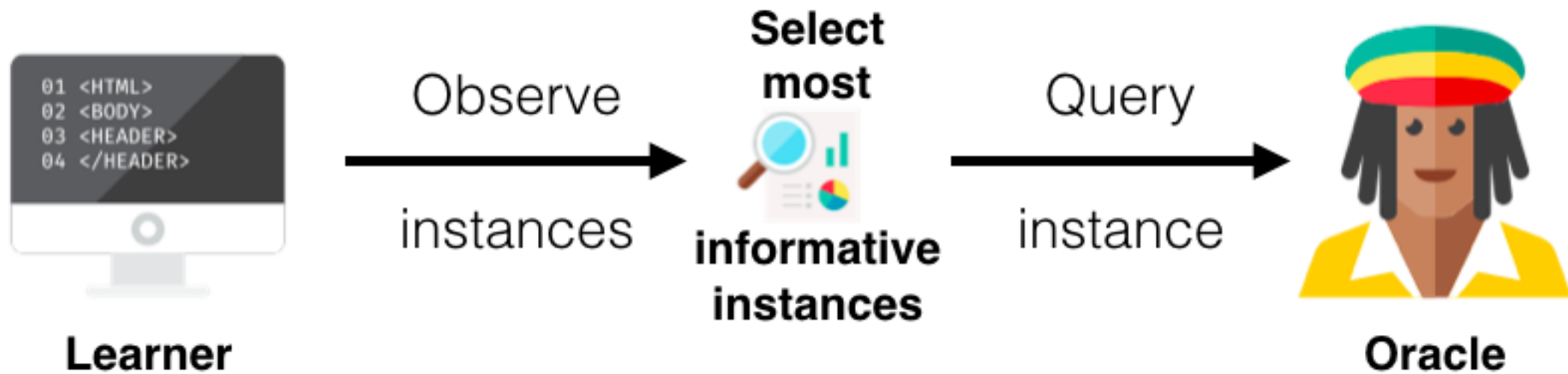
Stream-Based Selective Sampling



Membership Query Synthesis



Pool-based Sampling



Single Instance / Batch

Active Learning Methods

Uncertainty Sampling

Label points with the lowest confidence for the predicted class

Margin Sampling

Label points with the lowest difference between confidence of predicted class and 2nd best class

Entropy Sampling

Label points with highest entropy for predicted classes

Active Learning Methods

Query by Committee

A variety of models are trained on the current labeled data, and vote on the output for unlabeled data

Label those points for which the "committee" disagrees the most

Expected error reduction

Label those points that would most reduce the model's generalization error

Variance reduction

Label those points that would minimize output variance, which is one of the components of error

Active Learning Methods

Balance exploration and exploitation

Choose examples to label as a trade-off between exploration and exploitation of the data space

This strategy manages this compromise by modelling the active learning problem as a contextual bandit problem.

For example, Bouneffouf et al.[7] propose a sequential algorithm named Active Thompson Sampling (ATS), which, in each round, assigns a sampling distribution on the pool, samples one point from this distribution, and queries the oracle for this sample point label.

Today's Hands-on Agenda – Active Learning Part

- 2 Datasets –
 - Image - the MNIST digits dataset
 - Text – spam/ham text messages
- Train a simple Logistic Regression model
- Check if model confidence correlates with its performance
- Train several models (LR, KNN, SVM), and check when they agree/disagree
- (Finally!) Test it ourselves - can Active Learning improve our performance?

Today's Hands-on Agenda – Active Learning Part

(Finally!) Test it ourselves - can Active Learning improve our performance?

1. We set an initial batch for labeling
2. We then select N new examples and ask the Oracle for their true label
3. We train a model on examples with labels
4. Repeat steps 2-3



Omri Allouche



Tal Perry



Inbal Horev



Eran Marom



Yonathan Guttel

Using Unlabeled Data

"Believe in yourself" - Bootstrapping

- The number of labeled instances is often too small for proper classification
- Better performance are obtained when weighting labeled instances with unlabeled instances
- Run EM several times improves performance
- Biggest gain is achieved on the 1st run of EM - for performance, run EM only once

Active Learning Theory

Select instances according to

1. *Informativeness*
2. *Representativeness*
3. *Diversity*

How to select seed?

- Represent all classes
- Perhaps first cluster and select from different clusters
- Start with potentially difficult set (e.g. "I will" for Action Items)
- Consider class priors

Criticism against Active Learning

- Selected instances often tied to chosen algorithm
 - Do instances selected because of one model's confidence be helpful for a different model?
- Selected instances are a biased distribution of population
- Performance depends on the proficiency of annotator
 - Labelers tend to make mistakes for “difficult” instances

Proactive Learning

Set to relax the unrealistic assumptions of Active Learning

1. Oracle may be wrong
2. especially when question is more difficult
3. Oracle may be reluctant to answer
4. Cost of labeling increase for difficult questions

When Active Learning is a Bad Choice

1. Training takes very long
2. Labeling is very cheap
3. Classes are well-balanced
4. You are unsure regarding your choice of algorithm
5. There's no correspondence between model confidence and performance

Why Active Learning didn't Take Off

- In Academia –
 - Challenges include labeled instances and focus on overall performance, without considering the number of labels used
 - When creating your own dataset, you select instances that are “interesting”
- While labeling often requires only a simple share of a file, Active Learning requires an interactive platform
 - UI
 - Server
- Results aren't very robust (as our hands-on work shows), and depend on the algorithm
- There's no foolproof “Recipe” for Active Learning

Suggested Workflow

- Collect a small set of labeled instances using “Guided Search”
 - Use domain knowledge
 - Use statistics to propose candidates that a human can label
- Build a basic model, and evaluate performance
 - Learning Curve
 - Correlation between confidence and performance
 - Review manually instances with low confidence – there are often labeling mistakes there
- Decide where to invest your efforts
 - Fix errors with a rule-based model
 - Enlarge training data – use model confidence to highlight good candidates for labeling
 - Increase training set randomly
 - ...



We're hiring!

- Full-stack Java Engineers
- Machine-learning Engineers
- Data Scientists

Time for more
coffee!