

# Beyond Collaborative Filtering: ML & Recommendations at Meetup

Evan Estola  
Meetup.com  
evan@meetup.com  
@estola

# My Background

- Machine Learning Engineer
- At Meetup since May 2012
- BS Computer Science
  - Information Retrieval
  - Data Mining
  - Math
    - Linear Algebra
    - Graph Theory


All Meetups



within 25 miles of Los Angeles, CA

Groups

Calendar



## LA Robotics Club

**1,848 Robot Enthusiasts**


Next Meetup: Jun 15



## LA Machine Learning

**1,743 Members**


Next Meetup: Jun 16



## Future Chief Data Scientists in Orange County, CA


**219 Data Scientists**

Next Meetup: Tomorrow



## Open Data in Greater Los Angeles

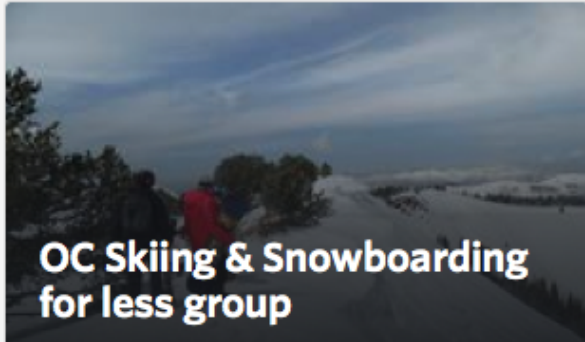
**246 open data users, data scient...**



## LA Arduino Enthusiasts

**395 Arduino Enthusiasts**

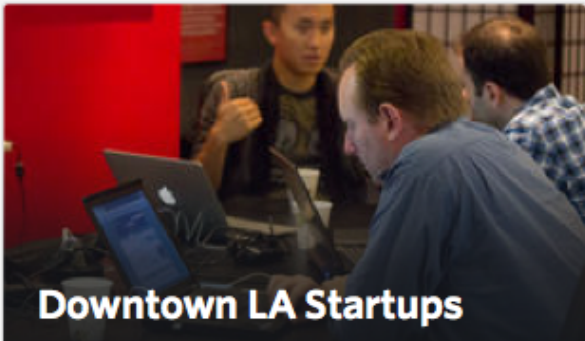
Next Meetup: Jun 26



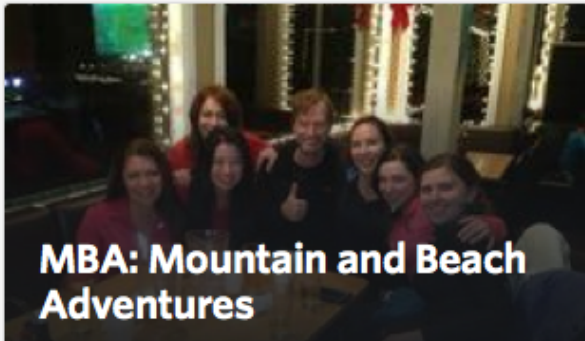
## OC Skiing & Snowboarding for less group

**1,537 Snowboarders & Skiers**


Next Meetup: Jun 28



## Downtown LA Startups



## MBA: Mountain and Beach Adventures

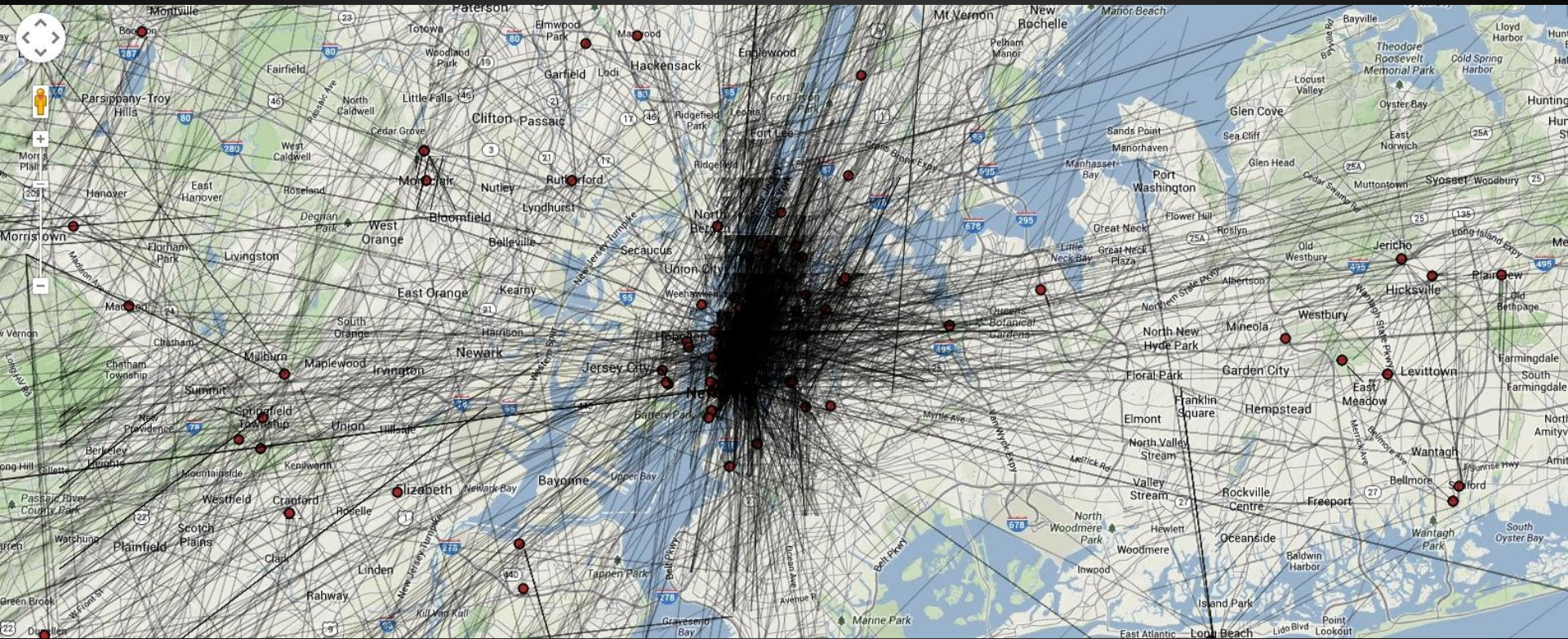


## Southern California MicroStrategy User Group

# Why Meetup data is cool

- Real people meeting up
- Every meetup could change someone's life
- No ads, just do the best thing
- Oh and >125 million rsmps by >17 million members
- ~3 million rsmps in the last 30 days
  - >1/second





# Tools at Meetup

- Hive - SQL on Hadoop
- Spark - Distributed Scala on Hadoop cluster
- Scala - Recommendations service
- R - Data analysis, Model building
- Python - Scripting, Data organizing
- Java - Backend of our web stack

# “Everything is a recommendation”

The screenshot shows the Meetup website interface. At the top, a red banner displays the Meetup logo and a personalized message: "Your calendar for the week of August 26" and "Hi Evan Testola, here's what Meetup looks like for you this week." Below this, a navigation bar includes links for "Find" and "Start" a Meetup Group, along with user options like "Evan Estola", "What's new", "Help", "My Groups", "Account", and "Log out". The main section is titled "NYC Data Science" and features a "Members" tab. The "Members" section shows a list of members, including "Chris Halpert" and "Randy Au", with details about their join dates and last visits. A sidebar on the left provides additional information about the group, such as "New York, NY", "Founded Jan 10, 2013", and a list of related topics like "Data Scientists", "Group reviews", "Past Meetups", and "Our calendar".

The screenshot shows a user profile for "Ian Schenck" from "New York, NY", who has been a member since July 26, 2010. The profile is part of a "Suggestions" section for the month of "Aug 2013". Below the profile, a list of recommended groups is shown, including "Go Language New York (aka golang ny)" with 395 members, "hackerschool & friends" with 23 members, and "Longboarding New York City" with 988 members. A link to "And 3 more groups (see all)" is provided at the bottom of the suggestions list.



## What are you interested in?

Arts & Culture ▾	Beliefs ▾	Business ▾	Crafts & Hobbies ▾												
Dancing ▾	Food & Drink ▾	Games & Sci-Fi ▾	Identity ▾												
Languages ▾	Literature ▾	Moms & Dads ▾	Movements ▾												
Music ▾	<b>Outdoors ▲</b>	Pets ▾	Photo & Films ▾												
<table><tr><td>+ Bicycling</td><td>+ International Travel</td><td>+ Hiking</td></tr><tr><td>+ Fishing</td><td>+ Outdoor Recreation</td><td>+ Motorcycle Riders</td></tr><tr><td>+ Road Cycling</td><td>+ Urban Exploration</td><td>+ Trail Running</td></tr><tr><td>+ Weekend Adventure &amp; Travel</td><td>+ Geocaching</td><td>+ Snowboarding</td></tr></table>				+ Bicycling	+ International Travel	+ Hiking	+ Fishing	+ Outdoor Recreation	+ Motorcycle Riders	+ Road Cycling	+ Urban Exploration	+ Trail Running	+ Weekend Adventure & Travel	+ Geocaching	+ Snowboarding
+ Bicycling	+ International Travel	+ Hiking													
+ Fishing	+ Outdoor Recreation	+ Motorcycle Riders													
+ Road Cycling	+ Urban Exploration	+ Trail Running													
+ Weekend Adventure & Travel	+ Geocaching	+ Snowboarding													
Social ▾	Sports & Fitness ▾	Tech ▾	Well-being ▾												

Next



# Collaborative Filtering

- Classic recommendations approach
- Users who like this also like this

## Customers Who Bought This Item Also Bought



Foundation and Empire  
(Foundation Novels)  
► Isaac Asimov  
★★★★☆ (859)  
Mass Market Paperback  
\$7.19



Second Foundation  
(Foundation Novels)  
► Isaac Asimov  
★★★★☆ (859)  
Paperback  
\$7.19



Prelude to Foundation  
(Foundation, Book 1)  
► Isaac Asimov  
★★★★☆ (157)  
Mass Market Paperback  
\$7.19



Forward the Foundation  
(Foundation Novels)  
► Isaac Asimov  
★★★★☆ (80)  
Mass Market Paperback  
\$7.19



Foundation's Edge  
(Foundation Novels)  
► Isaac Asimov  
★★★★☆ (94)  
Mass Market Paperback  
\$7.19



Foundation and Earth  
► Isaac Asimov  
★★★★☆ (126)  
Mass Market Paperback  
\$7.19

# Weaknesses of CF

- Sparsity
- Cold Start
- Coverage
- Diversity

# Why Recs at Meetup are hard

- Incomplete Data (topics)
- Cold start
- Asking user for data is hard
- Going to meetups is scary
- Sparsity
  - Location
  - Groups/person
  - Membership: 0.001%
  - Compare to Netflix: 1%

# Cleaning data

- Schenectady
- Beverly Hills
- Fake RSVP boosts (+100 guests!)
- Rsvp hogs



# Real data is gross

- Preprocessing is critical!
  - missing data
  - outliers
  - log scale
  - bucketing
  - sampling bias

# Supervised Learning/Classification

- “Inferring a function from labeled training data”
- Joined Meetup group/Didn't join Meetup group

# Ranking

- Membership  $\ll$  expected error rate
  - Sample to 50/50 join/no-join
- Model output label no longer explicitly true
- Use a classifier that gives you a useful output


All Meetups



within 25 miles of Los Angeles, CA

Groups

Calendar



## LA Snowboarders 0.976

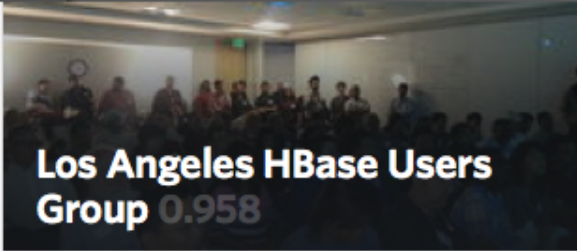
95 snowboarders

Next Meetup: Jul 19




## Let's Build A Mind Server 0.943

65 Inventors



## Los Angeles HBase Users Group 0.958

383 HBasers



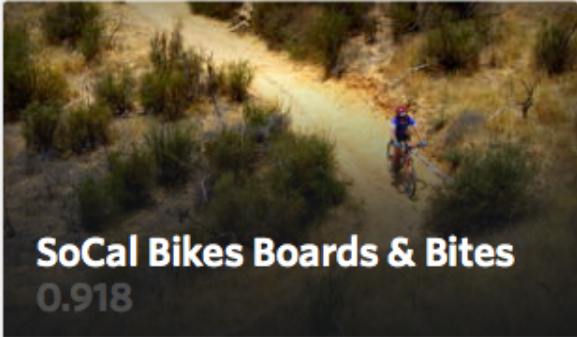
## Big Data Developers in Los Angeles 0.988

342 Members



## Los Angeles Racquetball Meetup 0.946


35 Racquetball players




## SoCal Bikes Boards & Bites 0.918

411 Riders


Next Meetup: Jul 5



## Los Angeles Hadoop Users Group- LA-HUG 0.976



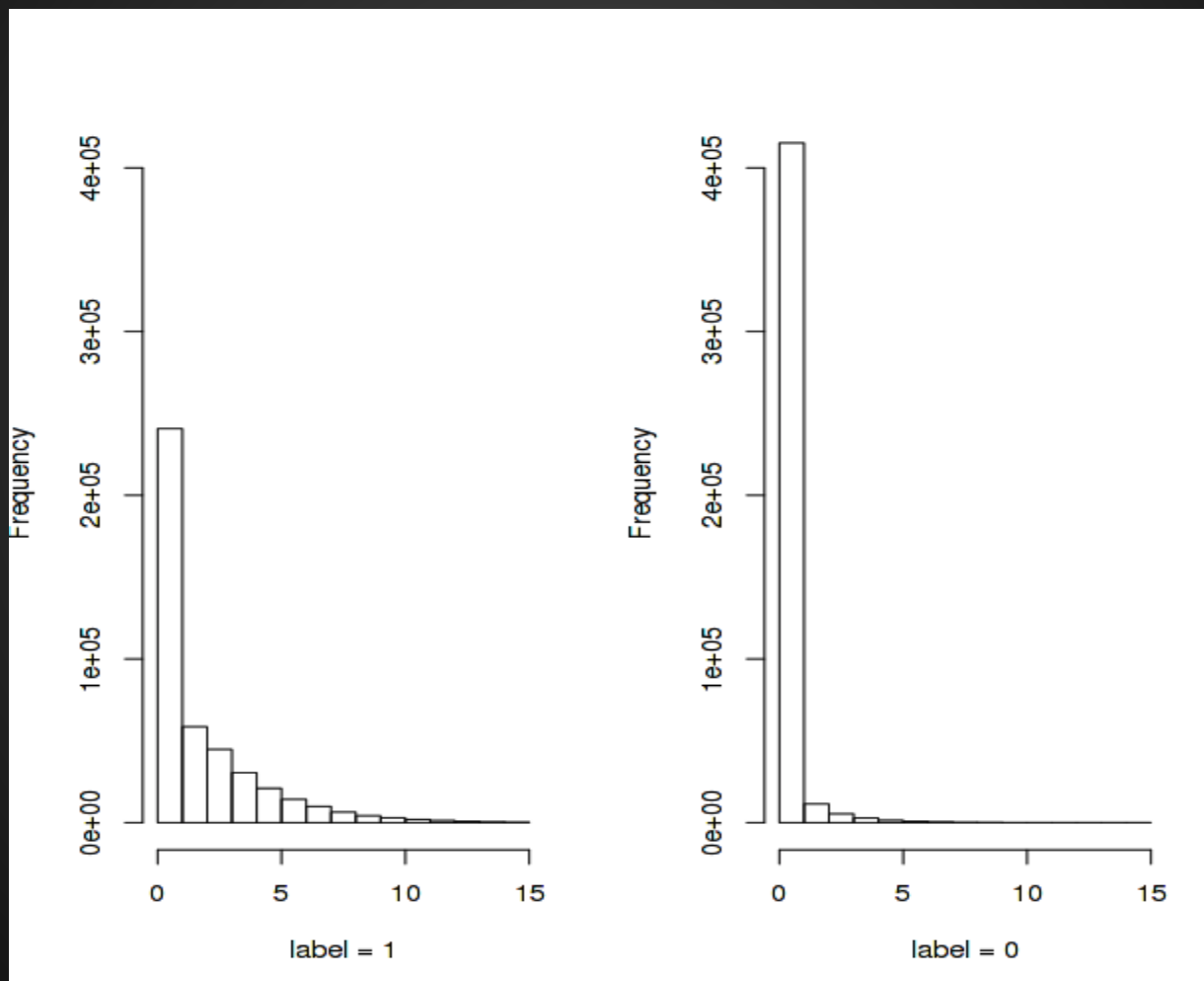
## Los Angeles Big Data Users Group 0.96



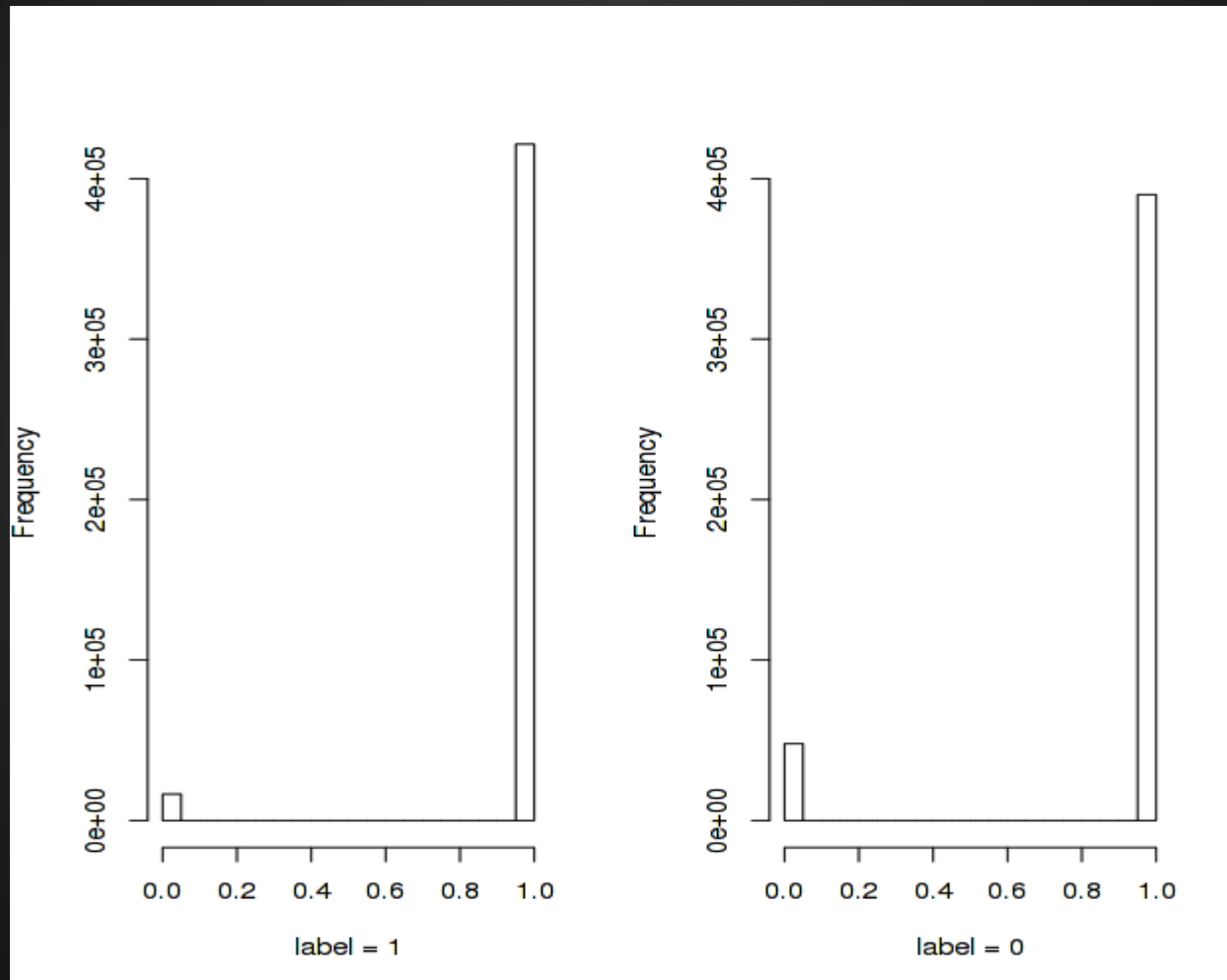
## LA Robotics Club 0.988



# Topic Match



# State Match



# Ensemble Learning

“... use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms”

# Ensemble Learning

- Collaborative Filtering on Topics
- Other simple features



# Logistic Regression Output

- RsvpScore 0.02
- FbFriends 2.02
- 2ndFbFriends 0.09
- AgeUnmatch -2.40
- GenUnmatch -3.37
- Distance -0.04
- StateMatch 0.54
- CountyMatch 0.41
- ZipScore 0.06
- TopicScore 4.14
- ExtendedTS 0.47
- RelatedTS 0.66
- FbLikeTS 0.78

# Facebook Likes

- Lots of information, but how to use?
- Map to topics, let training the model take care of the rest!
- Bonus: Recommendations server knows topics, generated topics can be passed in by request

# Mapping FB Likes to Meetup Topics

- Text based?
  - Go(game) vs Go(lang)?
  - Burton?
- Data approach!
  - Grab most popular topics across all members with the same like

# Normalization

- Top topics for Burton-Likers
  - Meeting New People, Coffee, bla bla
  - Most popular still dominates
- Normalize based on expected topic occurrence in sample

# Normalization

- For members with a given Like
- Compare percent with each topic to expected among total population
- Burton:
  - 20% “Meeting New People”
  - 9% “Snowboarding”
- Total population
  - 20% “Meeting New People”
  - 2% “Snowboarding”

# Processing

- Load FB Like connections, topics into Hadoop
- Process with Hive to generate top topics for each like
- Join with member likes to generate top topics per member
- Add feature to model using FB-Like-Generated-Topics crossover with groups...

# Results

- Positive weight
  - Very good sign
  - Captured information about member identity, not just behavior
- Deploy/Split test
  - 1.5% lift in conversion overall
  - (Only have facebook data for ~10% of members)



# Thanks!

Smart people come work with me.

<http://www.meetup.com/jobs/>