Meetup

Data Science Campinas

Discovering Users' Topics of Interest in Recommender Systems

Gabriel Moreira - @gspmoreira

ciandt.com

Lead Data Scientist

ITA

DSc. student

Recommender Systems

An introduction

Life is too short!

Social recommendations

Recommendations by interaction

"A lot of times, people don’t know what they want until you show it to them."

Steve Jobs

"*We are leaving the Information Age and entering the Recommendation Age.*".

Cris Anderson, "The long tail"

Recommendations are responsible for...

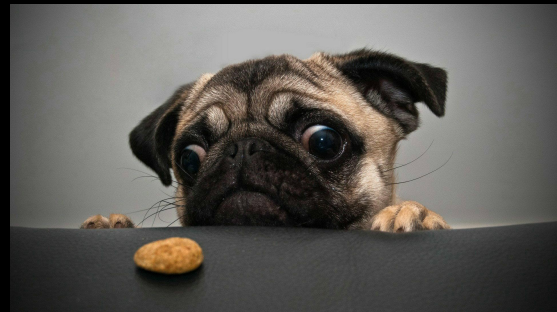
38% of sales

2/3 views

38% of top news visualization

What else may I recommend?

products
tags
professionals
courses
musics
movies
jobs
books
papers
girlfriends
investiments
restaurants
videos
dressing



What can a Recommender Systems do?

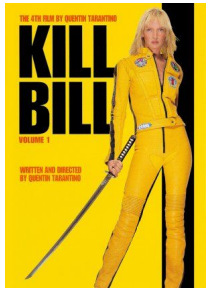
1 - Recommendation
Given a user, produce an ordered list matching the user needs

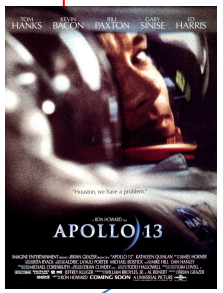
2 - Prediction
Given an item, what is its relevance for each user?

How it works

Content-Based Filtering

Similar content (e.g. actor)





Likes

Recommends

Content-Based Filtering

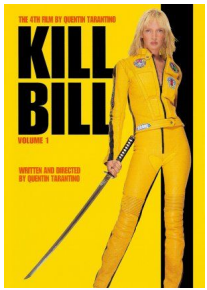
Advantages

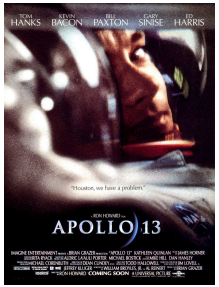
- Does not depend upon other users
- May recommend new and unpopular items
- Recommendations can be easily explained

Drawbacks

- Overspecialization
- May not recommend to new users
- May be difficult to extract attributes from audio, movies or images

User-Based Collaborative Filtering





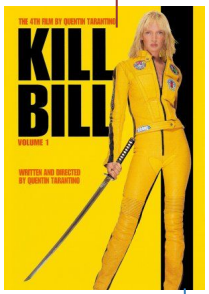
Likes

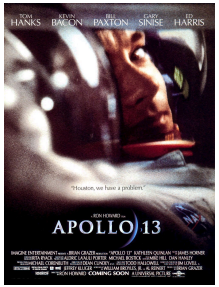
Recommends

Similar interests

Item-Based Collaborative Filtering

Who likes A also likes B





Likes

Likes

Likes

Recommends

Collaborative Filtering

Advantages

- Works to any item kind (ignore attributes)

Drawbacks

- Usually recommends more popular items
- Cold-start
  - Cannot recommend items not already rated/consumed
  - Needs a minimum amount of users to match similar users

Hybrid Recommender Systems

Some approaches

Composite
Iterates by a chain of algorithm, aggregating recommendations.

Weighted
Each algorithm has as a weight and the final recommendations are defined by weighted averages.

UBCF Example (Java / Mahout)

```
User,Item,Rating1,
15,4.0
1,16,5.0
1,17,1.0
1,18,5.0
2,10,1.0
2,11,2.0
2,15,5.0
2,16,4.5
2,17,1.0
2,18,5.0
3,11,2.5
```

input.csv

```
1 // Loads user-item ratings
  DataModel model = new FileDataModel(new File("input.csv"));
2 // Defines a similarity metric to compare users (Person's correlation coefficient)
  UserSimilarity similarity = new PearsonCorrelationSimilarity(model);
3 // Threshold the minimum similarity to consider two users similar
  UserNeighborhood neighborhood = new ThresholdUserNeighborhood(0.1, similarity,
  model);
4 // Create a User-Based Collaborative Filtering recommender
  UserBasedRecommender recommender = new
  GenericUserBasedRecommender(model, neighborhood, similarity);
5 // Return the top 3 recommendations for userId=2
  List recommendations = recommender.recommend(2, 3);
```

User-Based Collaborative Filtering example (Mahout)

https://mahout.apache.org/users/recommender/userbased-5-minutes.html

Frameworks - Recommender Systems

Java

Python / Scala

Python

Java

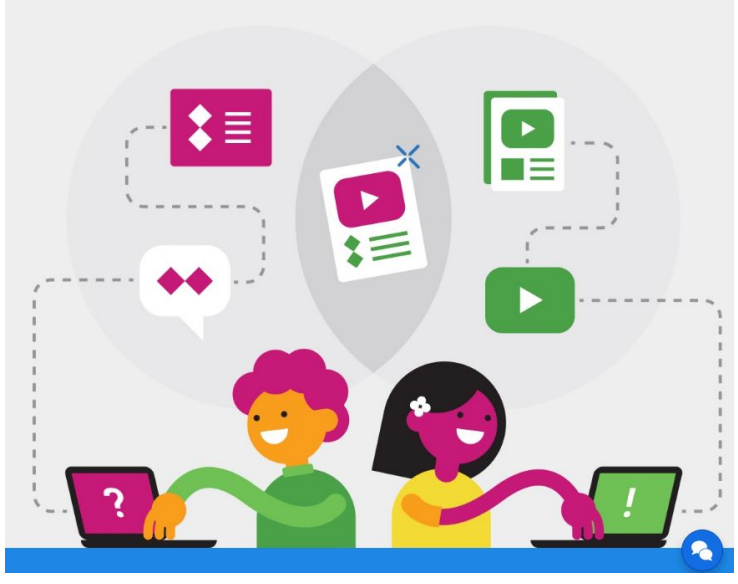
.NET

Case

Smart Canvas©

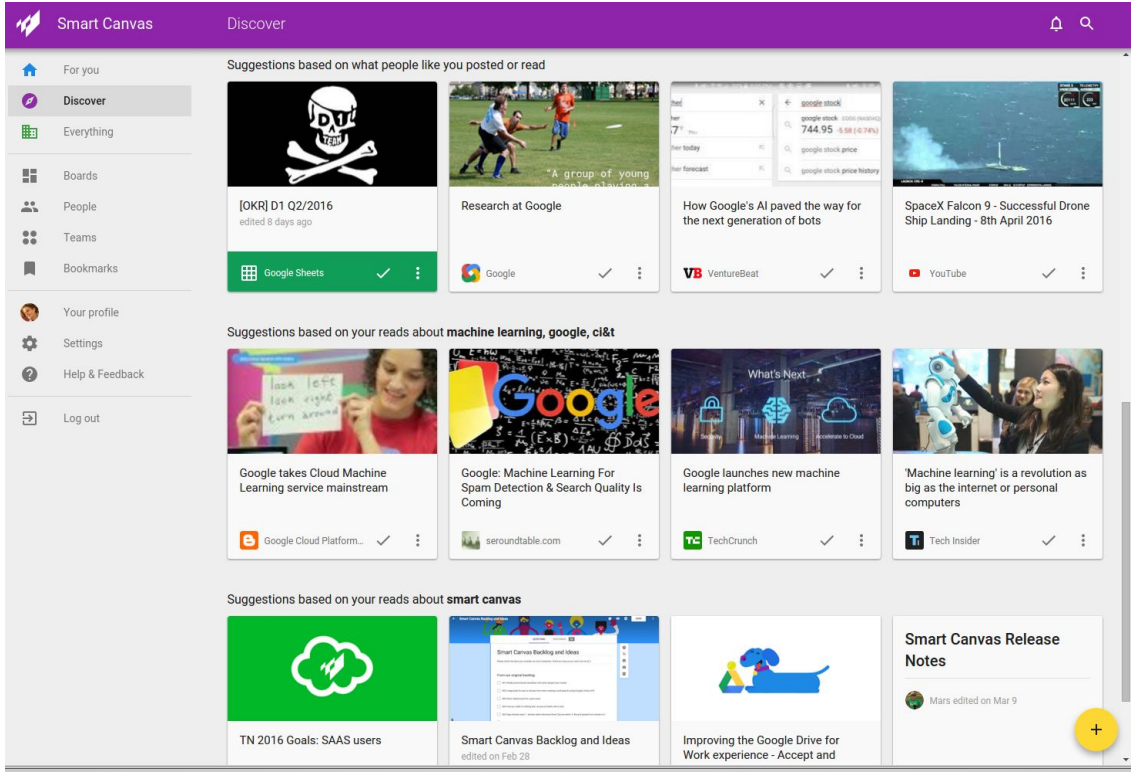
Corporate Collaboration

Powered by Recommender Systems and Topic Modeling techniques



http://www.smartcanvas.com/

Content recommendations

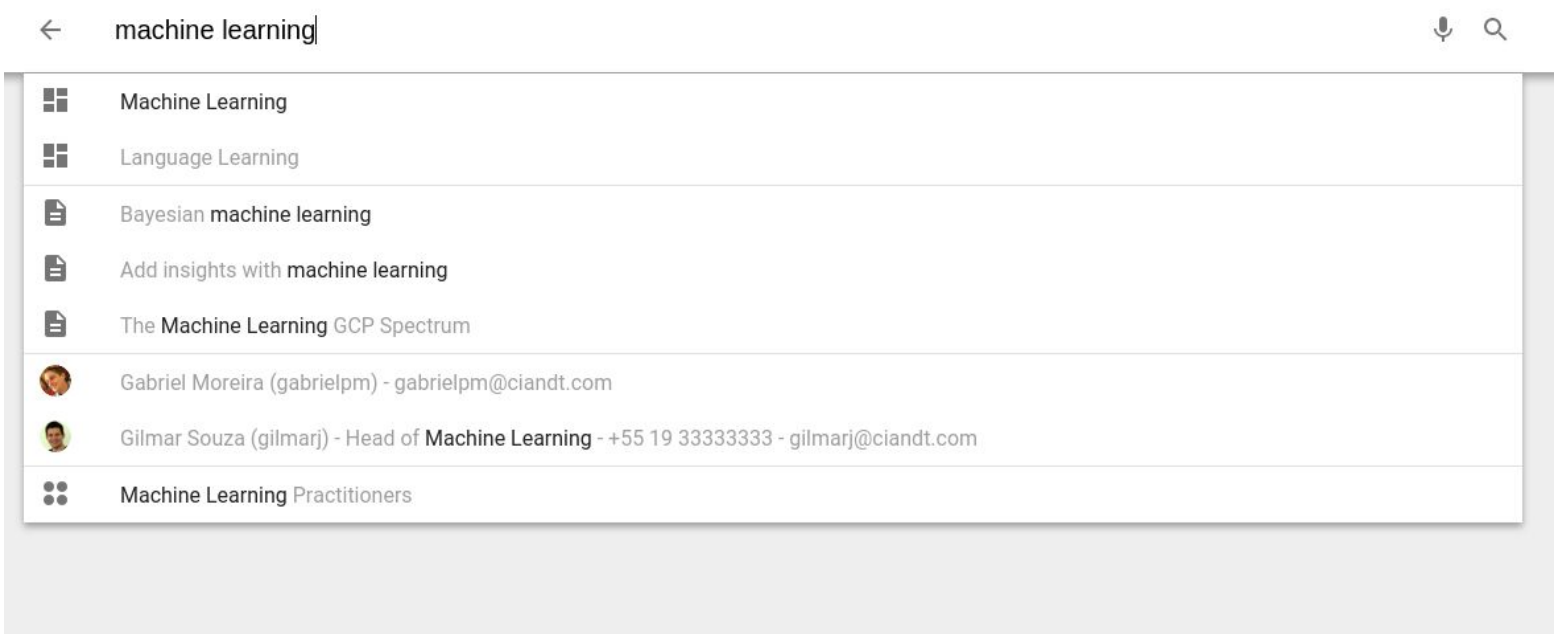


Discover channel - Content recommendations with personalized explanations, based on user's topics of interest (discovered from their contributed content and reads)

Person topics of interest

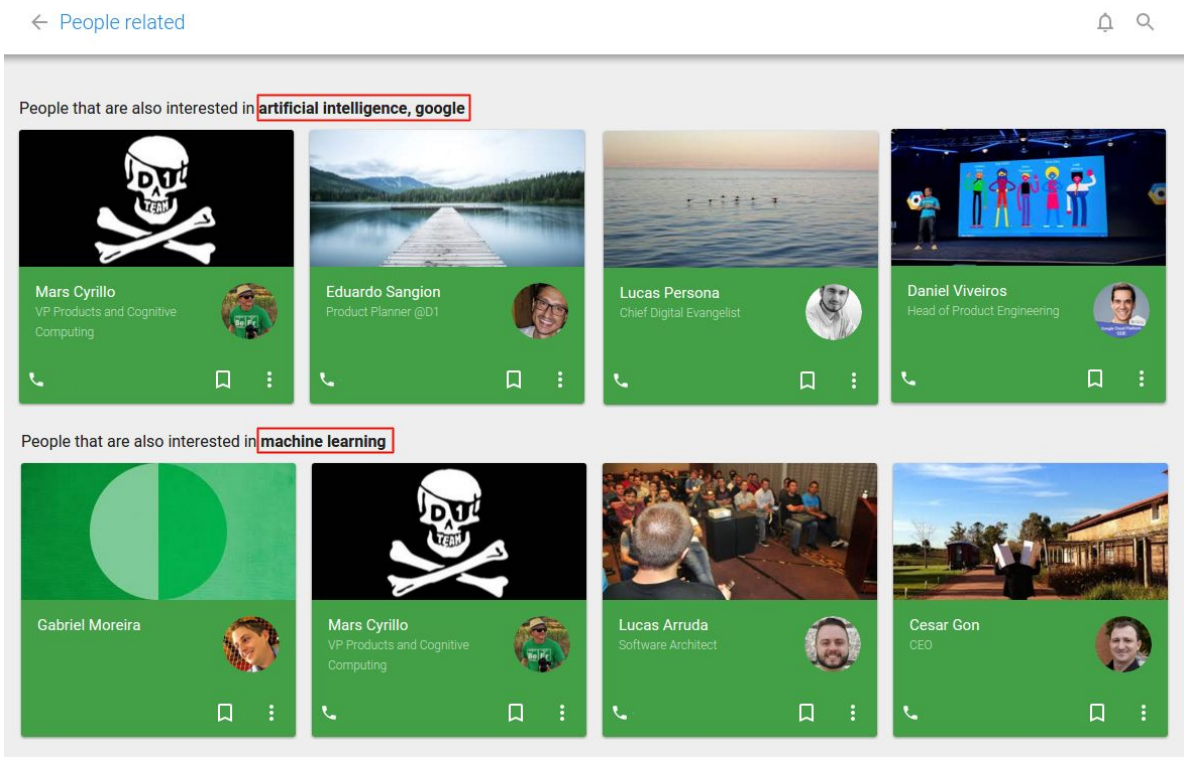
User profile - Topics of interest of users are from the content that they contribute and are presented as tags in their profile.

Searching people interested in topics / experts...



User discovered tags are searchable, allowing to find experts or people with specific interests.

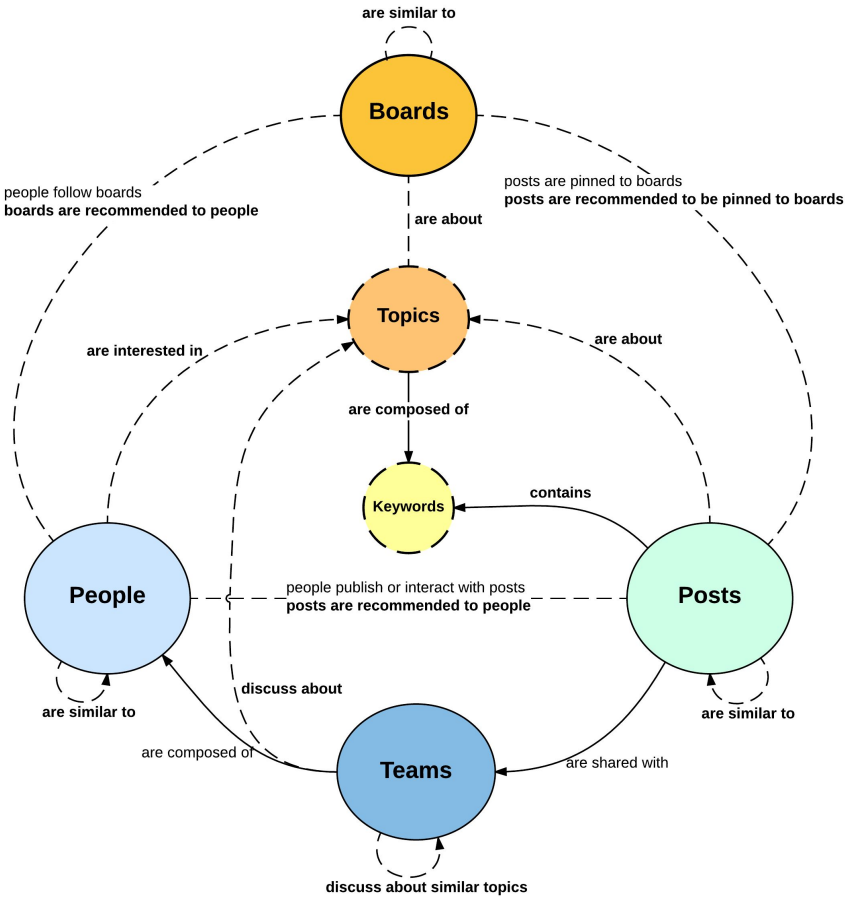
Similar people



People recommendation - Recommends people with similar interests, explaining which topics are shared.

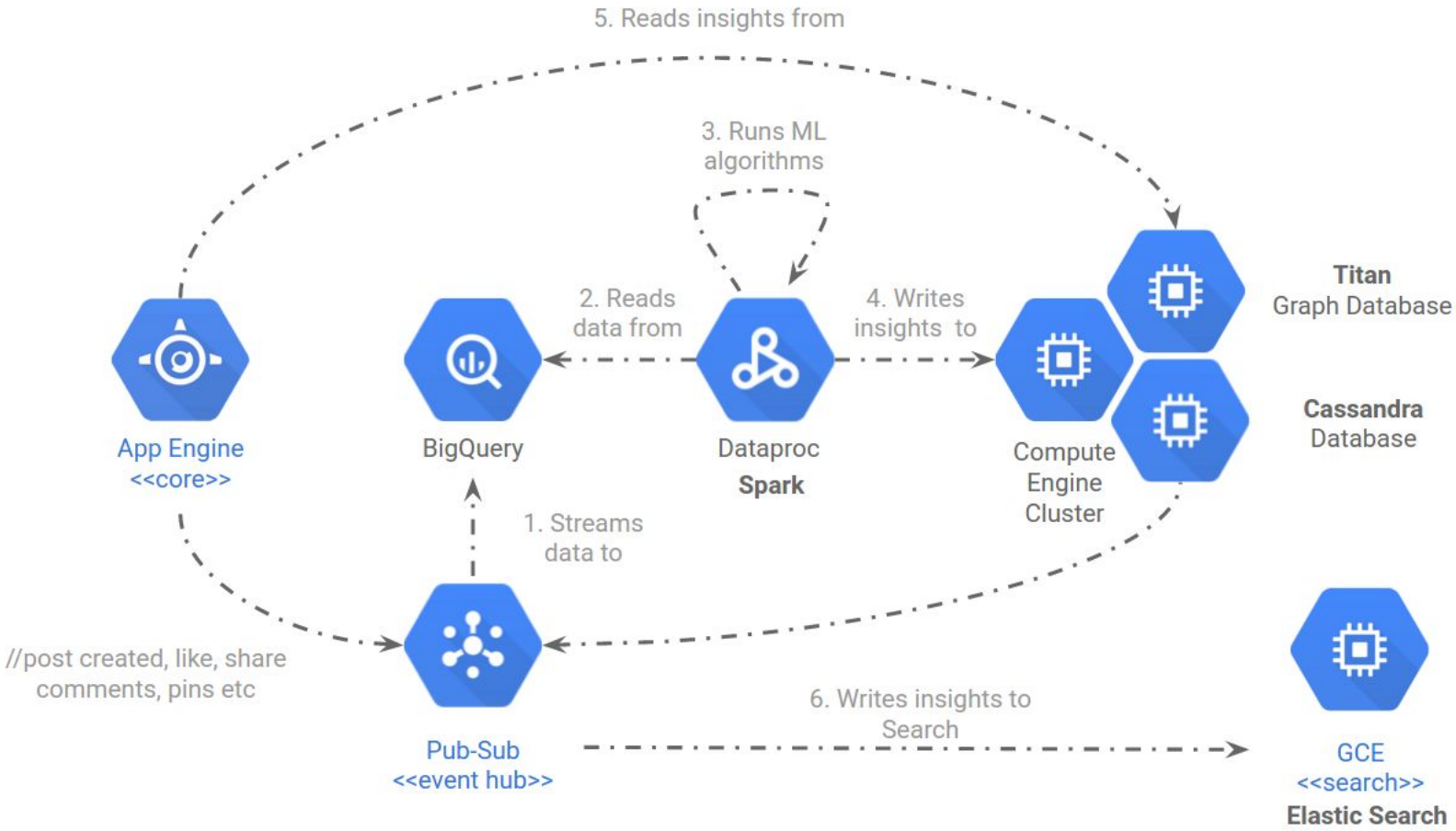
How it works

Collaboration Graph Model



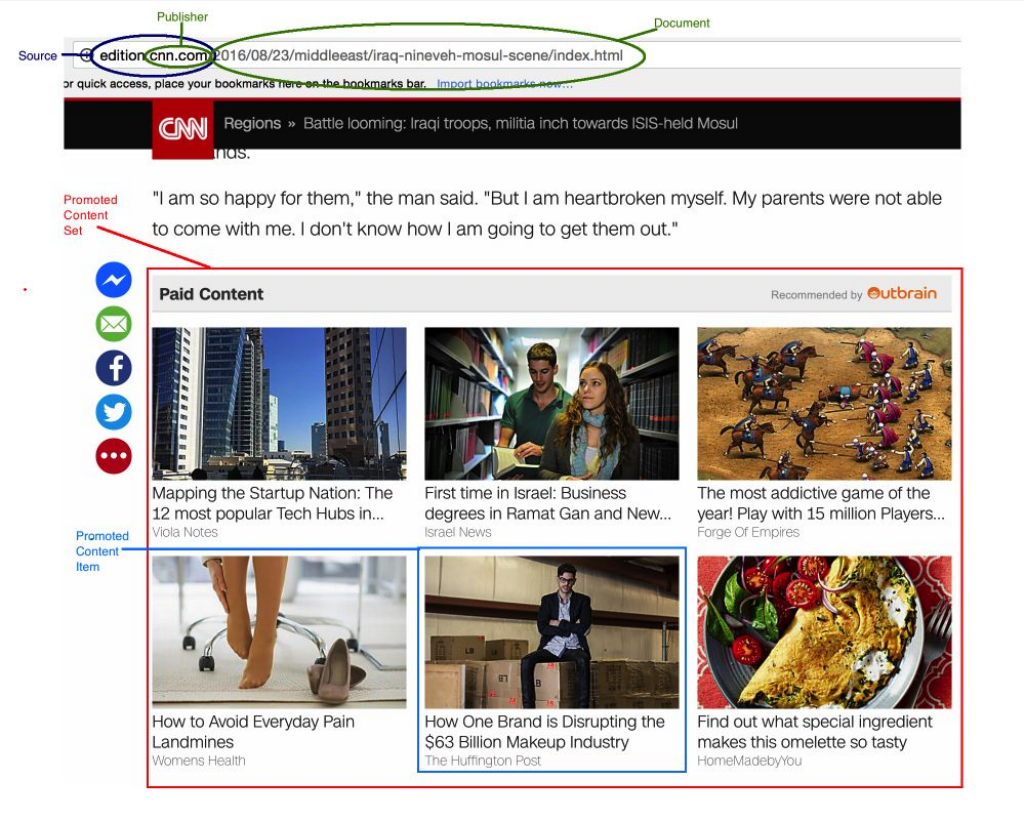
Smart Canvas © Graph Model: Dashed lines and bold labels are the relationships inferred by usage of RecSys and Topic Modeling techniques

Architecture Overview



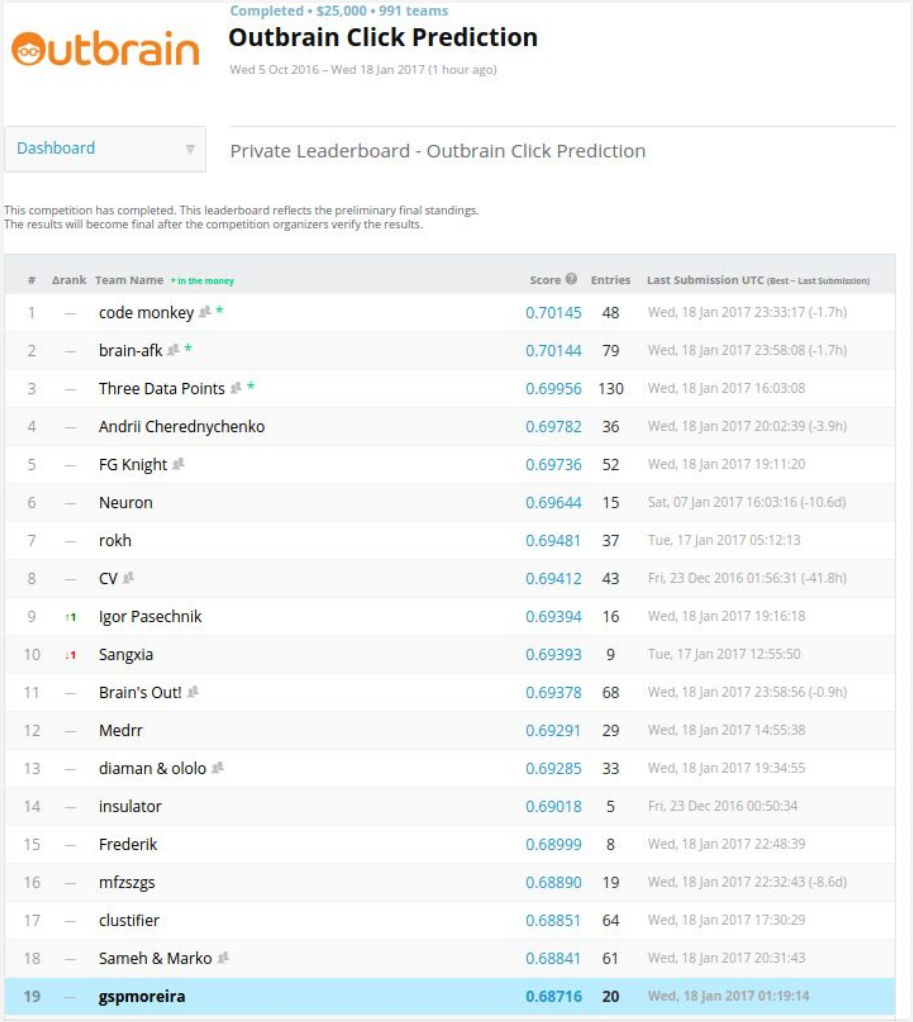
Outbrain Click Prediction - Kaggle competition

Can you predict which recommended content each user will click?



Dataset

- Sample of users page views and clicks during 14 days on June, 2016
- 2 Billion page views
- 17 million click records
- 700 Million unique users
- 560 sites



Completed • $25,000 • 991 teams

Outbrain Click Prediction

Wed 5 Oct 2016 – Wed 18 Jan 2017 (1 hour ago)

Dashboard

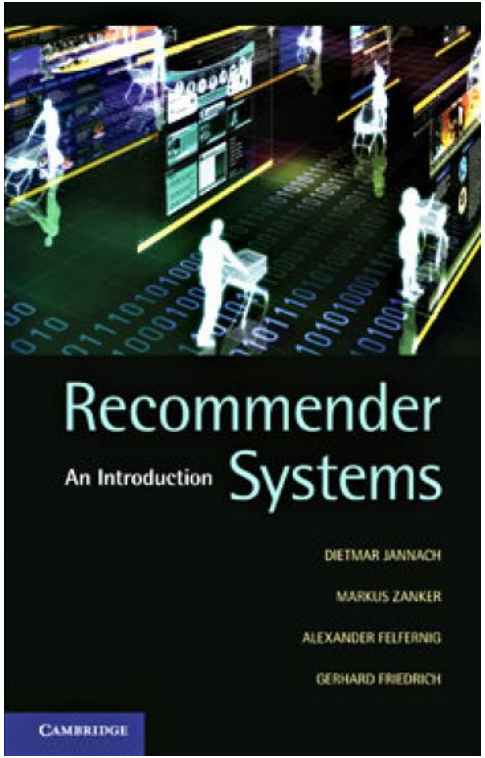
Private Leaderboard - Outbrain Click Prediction

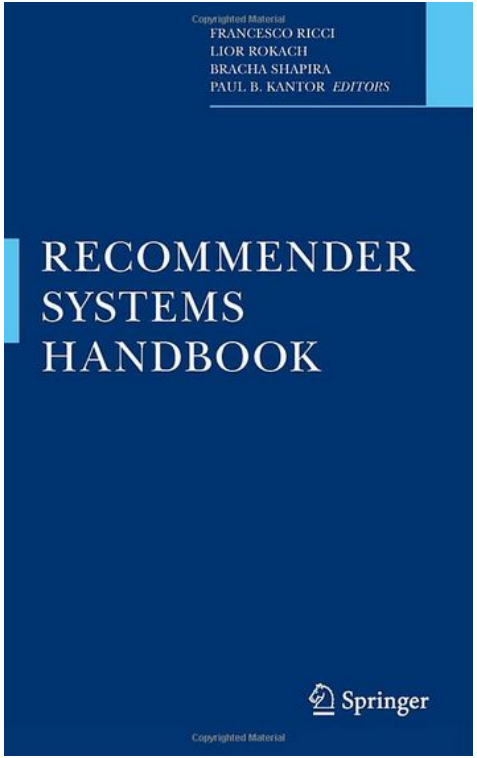
This competition has completed. This leaderboard reflects the preliminary final standings.
The results will become final after the competition organizers verify the results.

#	Δrank	Team Name * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	—	code monkey *	0.70145	48	Wed, 18 Jan 2017 23:33:17 (-1.7h)
2	—	brain-afk *	0.70144	79	Wed, 18 Jan 2017 23:58:08 (-1.7h)
3	—	Three Data Points *	0.69956	130	Wed, 18 Jan 2017 16:03:08
4	—	Andrii Cherednychenko	0.69782	36	Wed, 18 Jan 2017 20:02:39 (-3.9h)
5	—	FG Knight	0.69736	52	Wed, 18 Jan 2017 19:11:20
6	—	Neuron	0.69644	15	Sat, 07 Jan 2017 16:03:16 (-10.6d)
7	—	rokh	0.69481	37	Tue, 17 Jan 2017 05:12:13
8	—	CV	0.69412	43	Fri, 23 Dec 2016 01:56:31 (-41.8h)
9	↑1	Igor Pasechnik	0.69394	16	Wed, 18 Jan 2017 19:16:18
10	↓1	Sangxia	0.69393	9	Tue, 17 Jan 2017 12:55:50
11	—	Brain's Out!	0.69378	68	Wed, 18 Jan 2017 23:58:56 (-0.9h)
12	—	Medrr	0.69291	29	Wed, 18 Jan 2017 14:55:38
13	—	diaman & ololo	0.69285	33	Wed, 18 Jan 2017 19:34:55
14	—	insulator	0.69018	5	Fri, 23 Dec 2016 00:50:34
15	—	Frederik	0.68999	8	Wed, 18 Jan 2017 22:48:39
16	—	mfzszgs	0.68890	19	Wed, 18 Jan 2017 22:32:43 (-8.6d)
17	—	clustifier	0.68851	64	Wed, 18 Jan 2017 17:30:29
18	—	Sameh & Marko	0.68841	61	Wed, 18 Jan 2017 20:31:43
19	—	**gspmoreira**	**0.68716**	**20**	**Wed, 18 Jan 2017 01:19:14**

I got **19th** position
from about
1000 competitors
(top 2%),
mostly due to
Feature Engineering
techniques.

Books


Recommender Systems
An Introduction
DIETMAR JANNACH
MARKUS ZANKER
ALEXANDER FELFERNIG
GERHARD FRIEDRICH
CAMBRIDGE


Copyrighted Material
FRANCESCO RICCI
LIOR ROKACH
BRACHA SHAPIRA
PAUL B. KANTOR EDITORS
RECOMMENDER SYSTEMS HANDBOOK
Springer
Copyrighted Material

ciandt.com

Thanks!

bit.ly/recsys_ds_camp
bit.ly/kaggle_outbrain_fe

Gabriel Moreira
Lead Data Scientist
@gspmoreira