

Recommendation System

By Bianca Orozco



About Medium

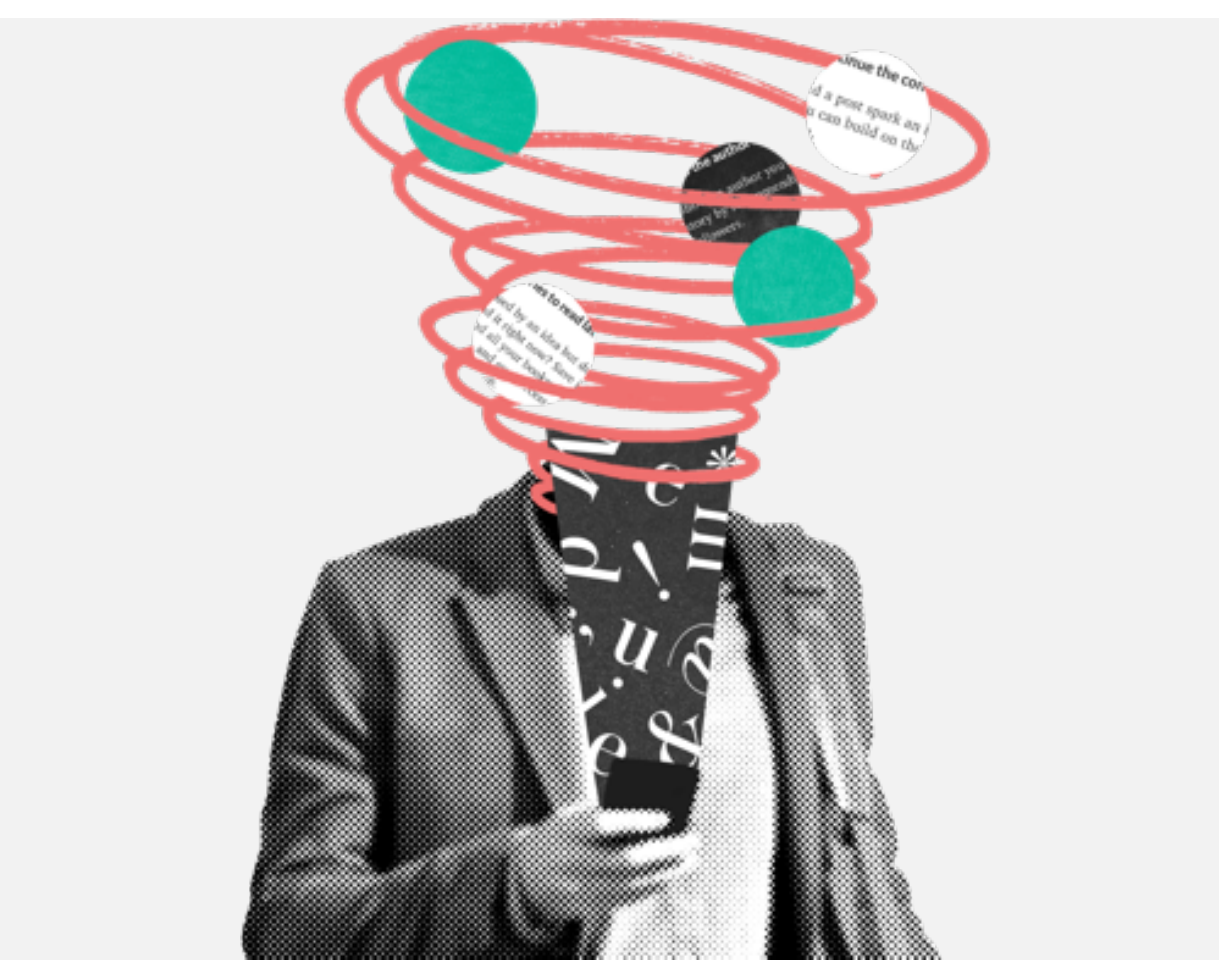


Facts:

- Online publishing platform (ad-free)
- Started in 2012 by Evan Williams
- 150 million monthly readers
- \$5 per month/\$50 yearly

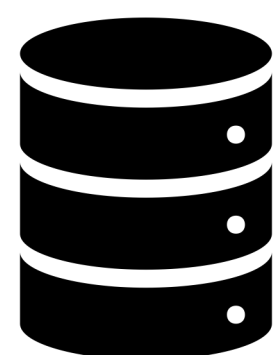
Popular Tags as of 2019:

1. Startup
2. Life
3. Politics
4. Life Lessons
5. Travel



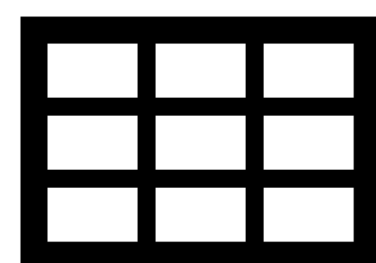


Methodons



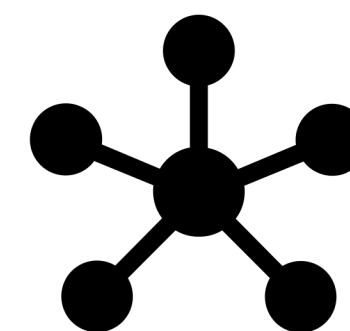
Data

- Kaggle
- 1.4 Million Titles
- 2017-2018
- 95 Tags



Cleaning/NLP

- Title + Subtitle
- TF-IDF
- Uni/Bigrams
- LSA



Clustering

- K-Means
- 4 Clusters



Recommendations

- Title Input
- Most popular related articles



Recommendation System

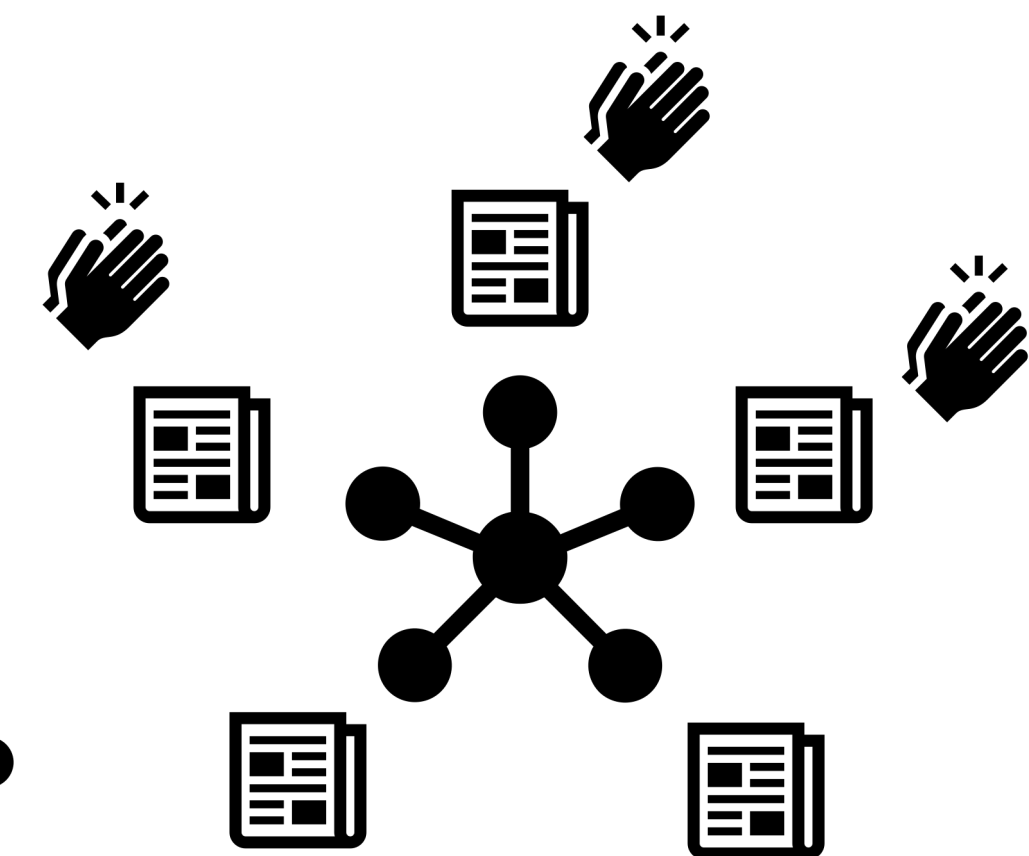


“Can Neural Networks Develop Attention?
Google Thinks They Can.”

TF-IDF

LSA

K-Means



Top 3 Related Articles:

1. How I Fully Quit Google (And You Can, Too)
2. When They Leave
3. Popping The Bubble: Cryptocurrency vs. Dot Com



Future Work



- Obtain full articles
- Use Cosine Similarity to find the closest articles, then the most popular
- Use a computer that can handle more data
- Optimize parameters (GridSearchCV)
- Update data (2019)
- Produce better results than a search bar



Questions / Comments

M



Appendix

	Title	Subtitle	Claps	Tags	Text
0	Beta Access	We are moving forward with final preparations ...	310	Tag_computer-science	Beta Access We are moving forward with final p...
1	MX Player For PC/Laptop Windows 8.1/8/7/10 Dow...	MX Player is your primary media player on your...	0	Tag_books	MX Player For PC/Laptop Windows 8.1/8/7/10 Dow...
2	Whats Your Author Horoscope?	Forget the goats and scorpionsweve reassigned ...	2200	Tag_psychology	Whats Your Author Horoscope? Forget the goats ...
3	International Womens Day:	Meet the Young Champion for Women Around the W...	0	Tag_football	International Womens Day: Meet the Young Champ...
4	Workplace Learning ProgramsWhere Will It Lead ...	Are you running a business or a nonprofit orga...	0	Tag_blockchain	Workplace Learning ProgramsWhere Will It Lead ...

Appendix

Topic 0

new, time, people, like, world, best, know, life, day, make

Topic 1

com, https, tinyurl, tinyurl com, https tinyurl, pdf, pdf https, read, www, online

Topic 2

new, blockchain, data, bitcoin, ico, technology, platform, token, android, crypto

Topic 3

android, app, best, data, business, download, google, like, free, apk

Appendix

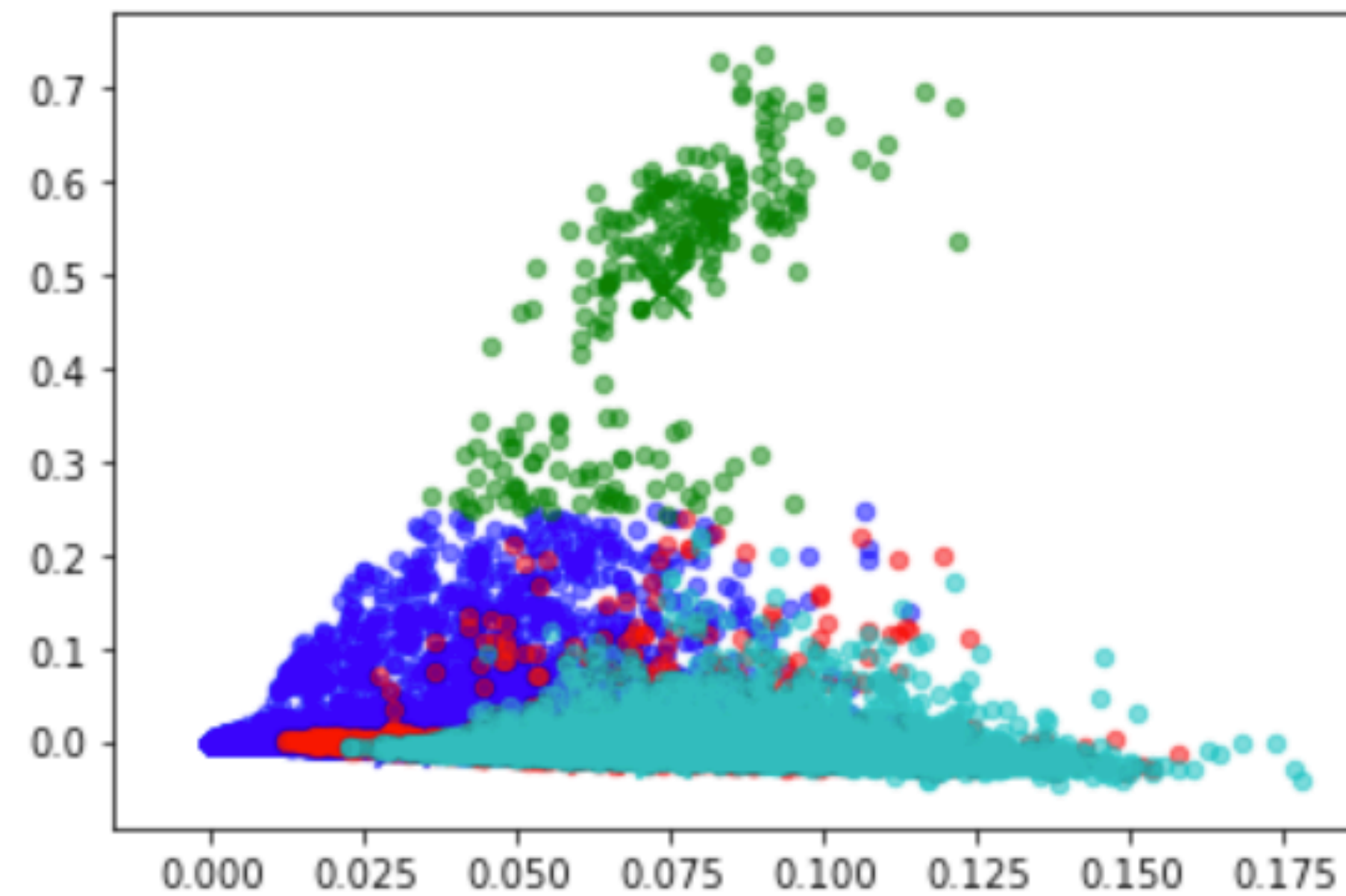
```
1 ▼ ##### Pipeline #####
2 ▼ pipeline = Pipeline([
3 ▼     ('tfidf', TfidfVectorizer(ngram_range=(1,2),
4                               binary=True,
5                               stop_words='english',
6                               min_df = 3,
7                               max_df = 0.75)),
8     ('lsa', TruncatedSVD(4)),
9 ▼     ('km', KMeans(n_clusters=4,
10                   init='k-means++',
11                   random_state=33))
12 ])
13
14 pipeline.fit_transform(docs)
15 pipeline.predict(docs) # Just to see what cluster they were assigned
```

executed in 6.86s, finished 09:22:00 2019-11-15

```
array([1, 1, 1, ..., 1, 1, 3], dtype=int32)
```

Appendix

Number of Documents: 42841



User Input: Can Neural Networks Develop Attention? Google Thinks They Can.
Inertia: 100.2806429526275