Soundcloud Spam Analysis

Pedro M. Sosa

Armin Ak

Metehan Ozten

Problem & Motivation

- Experiment with Machine Learning for Spam Analysis/Sybil Detection
- Unstudied Platforms: Soundcloud is a unique platform with 40 million users & limited ways for users to interact.
- Results relevant for similar platforms.
- Using less data than previous solutions

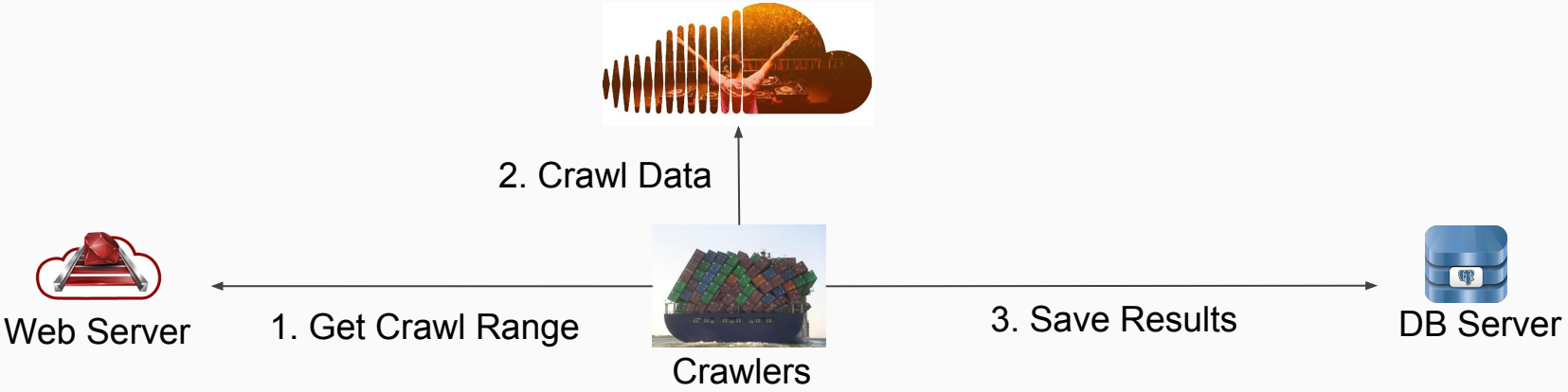
Data Mining

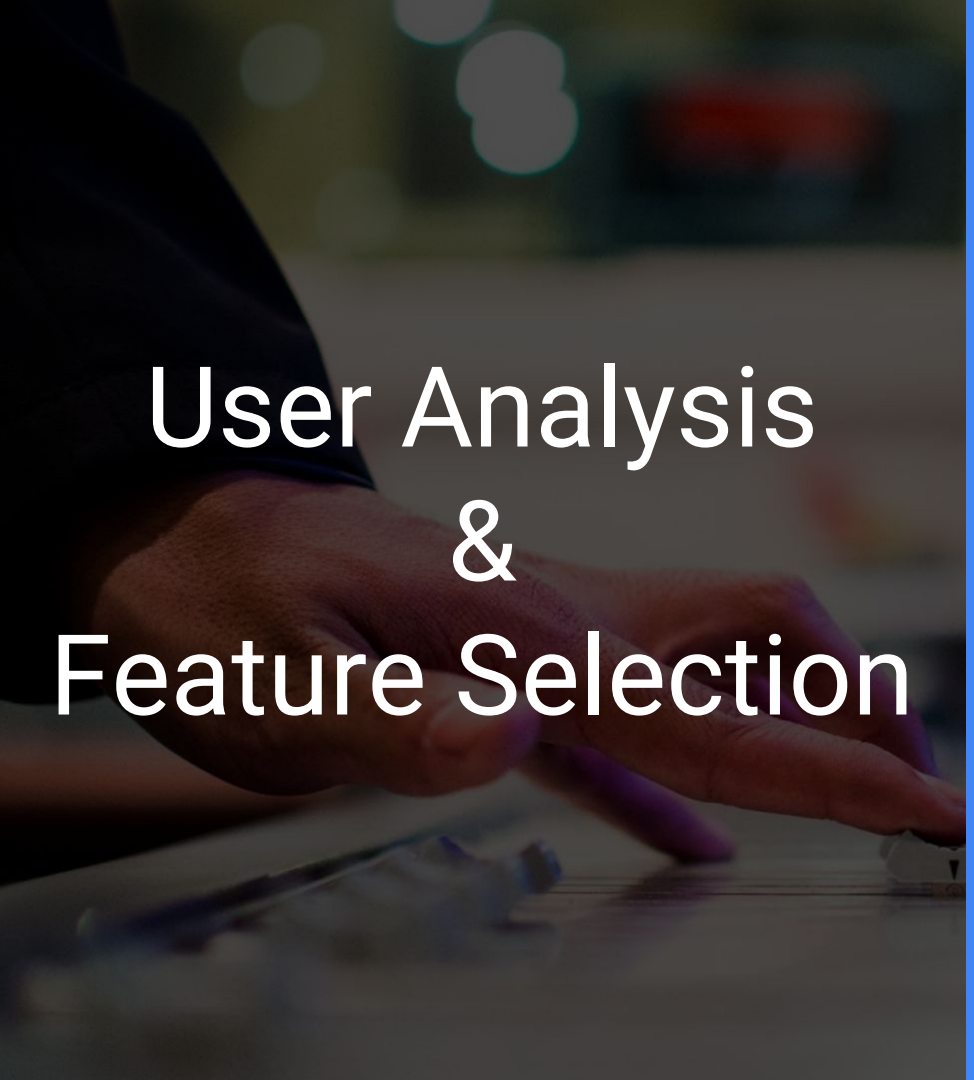
“Mine all the things!”

Data Mining

Dockerized distributed Crawler capable of pulling 100K users/hour

- Sequential IDs
- Example Call: *http://api.soundcloud.com/users/{ID}?client_id=**XXX***





User Analysis & Feature Selection

“Visualizing the Data”

User Analysis: User Engagement

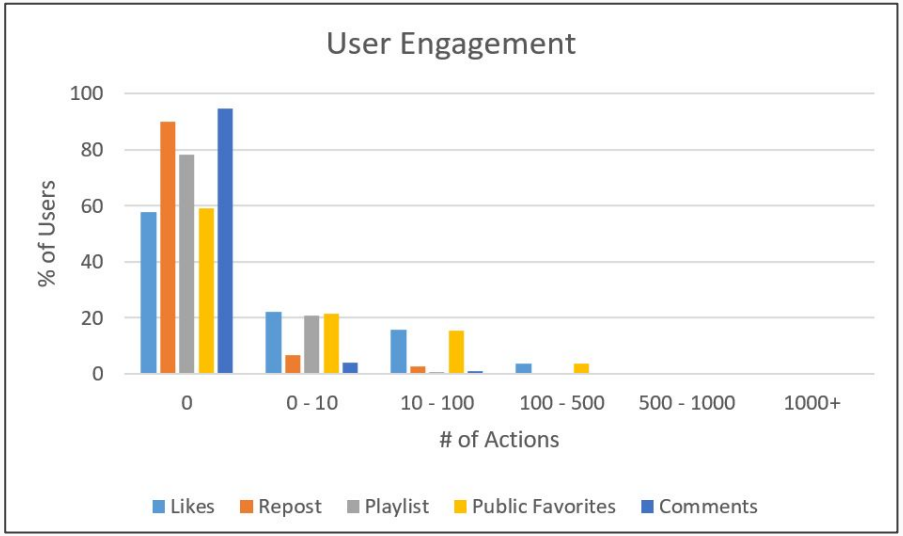
Subscriptions

Plan	% of Users
Free	99.79789337
Pro Plus	0.141886243
Pro	0.060214203
Solo	6.18025E-06

Users Content Generation

Track Count	% of People
0	92.81043084
0 - 10	6.018852134
10 - 100	1.153597016
100 - 500	0.015094993
500 - 1000	0.001068071
1000+	0.001136812

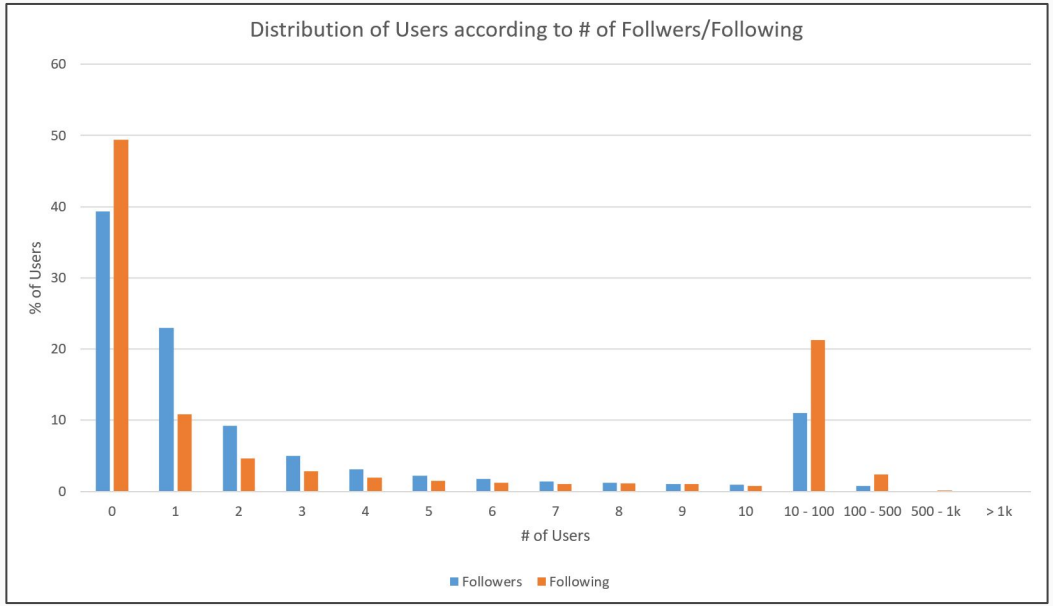
How do users engage with SoundCloud?

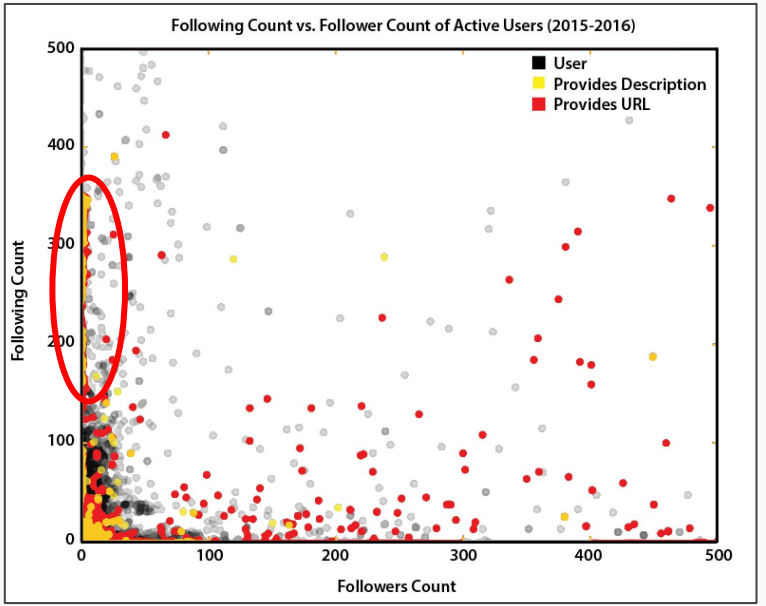


So what are users doing?

- 7.2% → Making Music
- 42.0% → Zombies
- 51.0% → Engage in **varying degrees.**

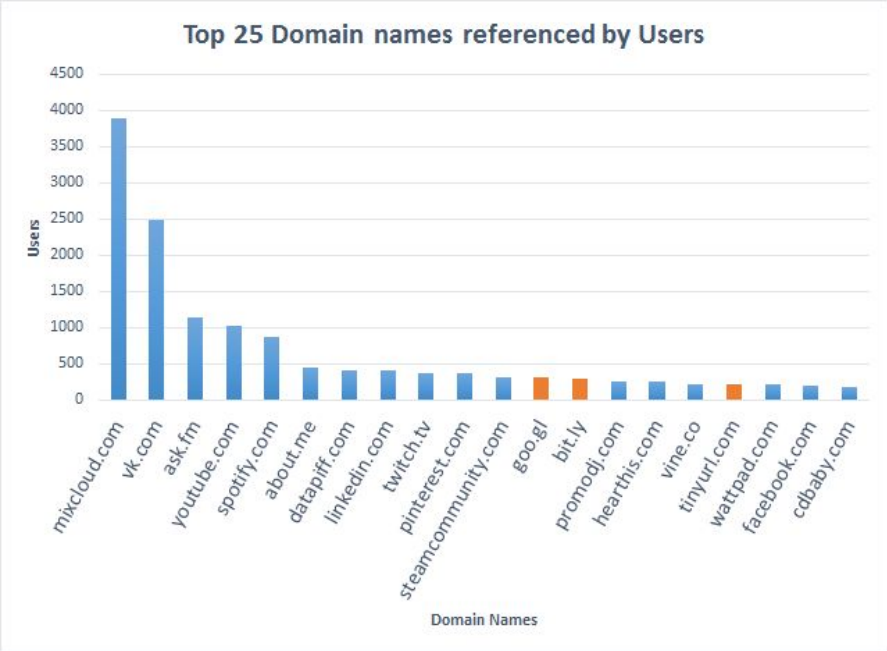
User Analysis: Following vs. Followers





User Analysis: Sharing Links

- 0.68% of users share URLs
- 0.004% use URL shorteners



Feature Selection

Feature	Description
Followers	of Followers
Following	of Following
Published	of Tracks Published
URL	T/F Do they provide a link?
ProfaneDesc	of Profanities in Description
ProfaneWeb	of Profanities in Website Title
Action	Likes + Reposts + Playlists + Favoirites + Comments
Pays	Do they pay for Soundcloud Pro/Go
Duplicate Description	of users that have the same description
Duplicate Website	of users that have that link to the same website
ShortenURL	Does their link point to a URL shortener service (e.g. bit.ly)

K-Means Clustering

“Life is like a box of clusters”

K-Means Clustering

Pre-processing Features

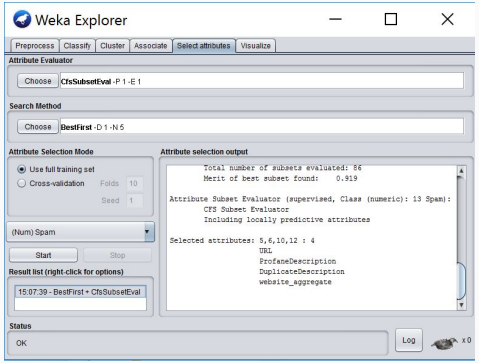
Method 1: Cluster Analysis

- Features consistently choosing centroids at 0

```
Centroid 0: [-0.00766839 -0.0443925  -0.01302298 -0.06668816 -0.03956699  0.
 -0.01321945  0.         -0.11812604  0.         -0.03965535]
Centroid 1: [ -8.46050088e-03   9.95456262e+00  -3.13315570e-02   1.49951660e+01
   2.52735922e+01   0.00000000e+00  -2.07116311e-01   0.00000000e+00
  -1.18407293e-01   0.00000000e+00   2.52172780e+01]
Centroid 2: [-0.00849098 -0.046821   -0.03133156 -0.06668816 -0.03956699  0.
  0.45036485  0.         8.45959806  0.         -0.03965535]
Centroid 3: [  0.47571996  20.62951992   1.33490661   0.66803644  -0.03956699   0.
   4.75573278   0.         -0.11840729   0.         -0.03965535]
Centroid 4: [  0.95160757   1.02779111   3.1285321   14.99516602  -0.03956699   0.
   0.79421633   0.         -0.07645566   0.          0.68197132]
[587, 891, 7897, 0, 12]
Centroid Buckets: [562100, 891, 7897, 719, 1613]
```

Method 2: Correlation-based Feature Subset Selection

- M. A. Hall (1998)
- WEKA



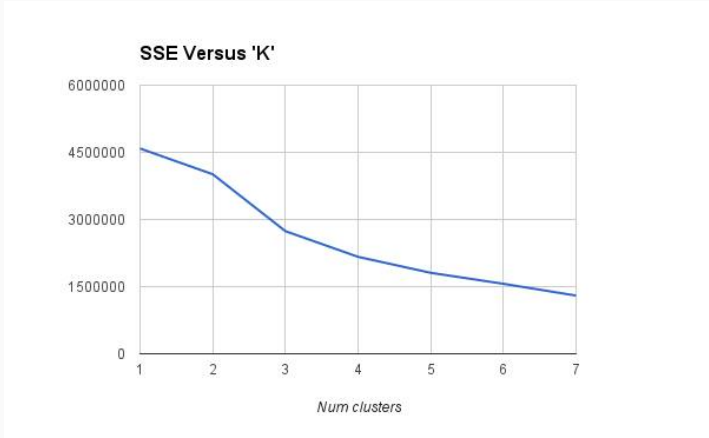
Method 3: Mean Removal & Z-Scaling

K-Means Clustering

Choosing # of Clusters

Method 1: Elbow Method

- Pick k to be either of the two values following the biggest drop in SSE for all clusters



Method 2: Natural Clustering

- Pick k to represent the natural amount of clusters
- 5 Clusters: Regular *Users, Popular Spammers, Unpopular Spammers, Celebrities*.

K-Means Clustering

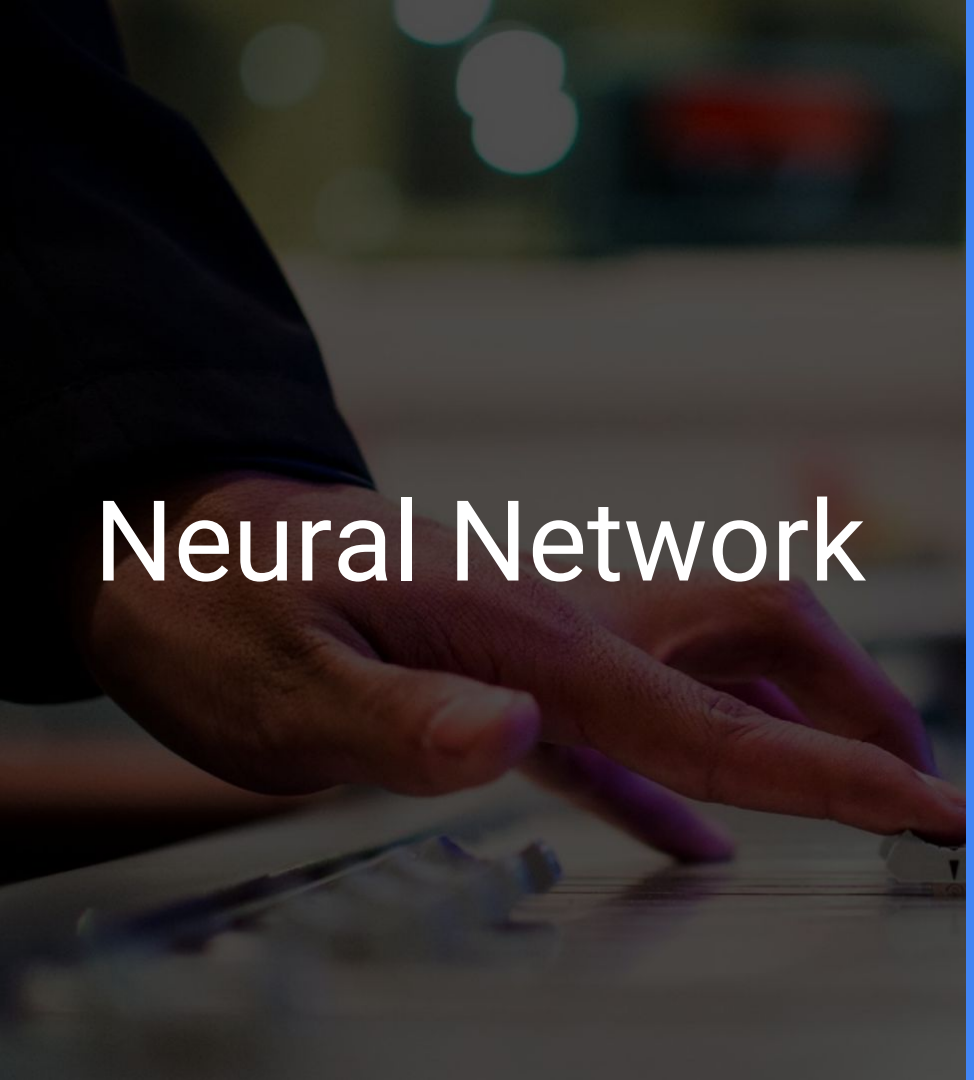
Results

K = 4

Cluster	Total Size	Spam	%
Regular Users	564319	587	<1%
Popular Spammers	902	902	100%
Unpopular Spammers	7987	7987	100%
Celebrities	102	1	<1%

K = 5

Cluster	Total Size	Spam	%
Regular Users	562121	587	<1%
Popular Spammers	891	891	100%
Unpopular Spammers	7987	7987	100%
Celebrities	1620	12	<1%
Super Fans	691	0	0%



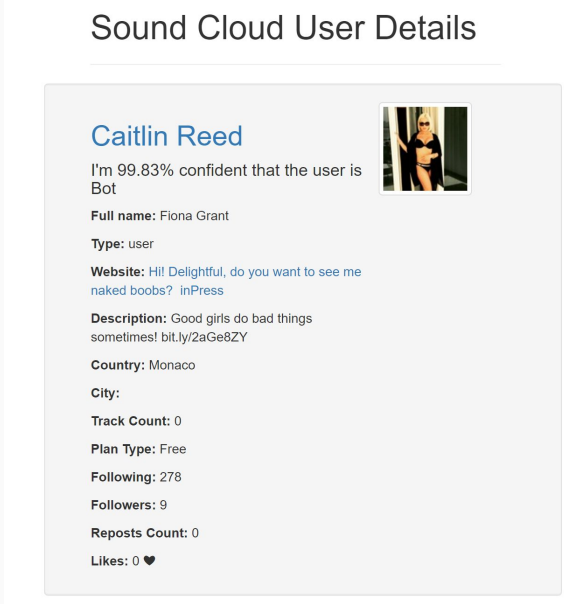
Neural Network

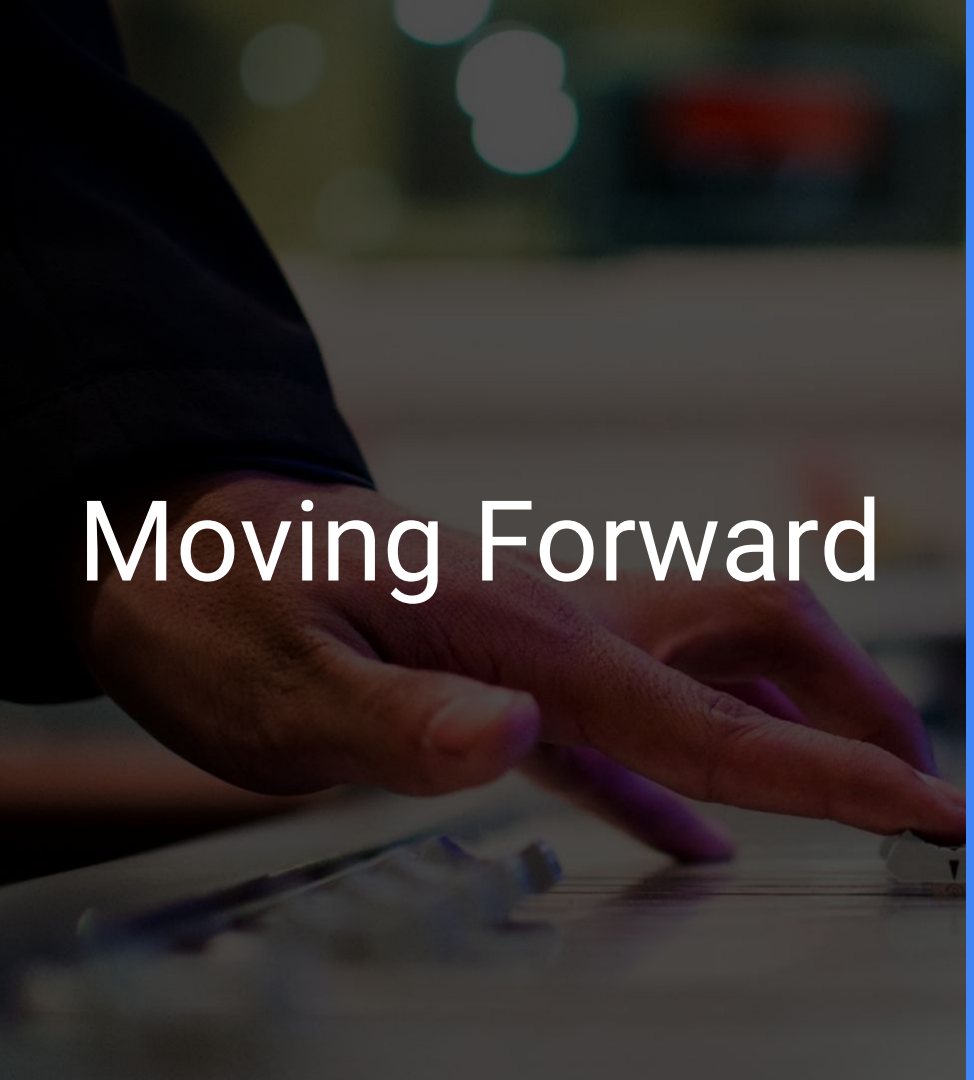
“I’m tired, let the machine do the work”

Neural Network

- Python's Keras Library
- Same features as with K-Means
- Specs:
  - Feed Forward
  - 1 Layer (ReLU)
  - 250 Nodes per Layer
- Train Data: 10,000 users
  - Musicians, Fans & Spam
- Test Accuracy: 85%

Demo at Soundcloud.pw





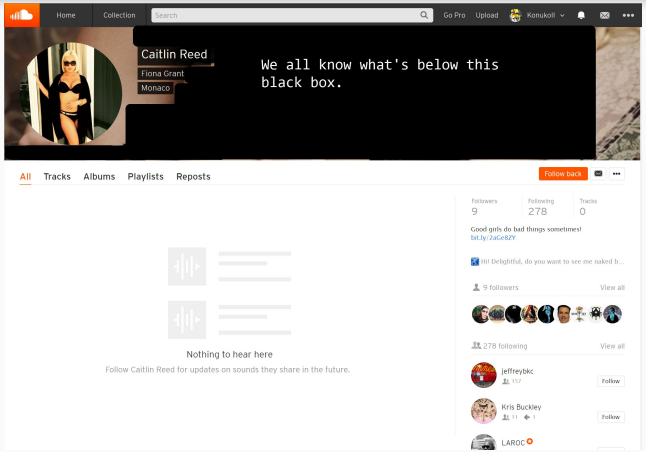
Moving Forward

“This is just the beginning”

Types of Spam and Origin

Pornographic Spam

- Use URL Shorteners to link to porn sites
- Profane Descriptions
- Re-used imagery and names



"Fake" Spam

- Very similar to legitimate users
- Descriptions based on randomly choosing simple phrases
- Created in batches
- Inactive?

"u'Allan Johnson'"	"u'Troublemaker. Student. Passionate food nerd. Coffee geek. Social media evangelist. Twitter specialist. Proud analyst.'"	'[]'	"u'user'"	"u'2016/10/08 15:39:22 +0
"u'Linda Neely'"	"u'Troublemaker. Student. Passionate food nerd. Coffee geek. Social media evangelist. Twitter specialist. Proud analyst.'"	'[]'	"u'user'"	"u'2016/10/08 15:39:22 +0
"u'Stephanie Brett'"	"u'Incurable music evangelist. Certified thinker. Evil entrepreneur. Social mediaholic. Proud analyst.'"	'[]'	"u'user'"	"u'2016/10/08 13:52:58 +0

Future Work

1. Continual On-the-spot Training of both the NN and K-Means.

2. Define more/better features.
   - Possibly make use of follower/following interactions

3. Follow up on “Fake Users” and “SEO leaches”. What do they do?

4. Extending our Crawler

Conclusions

Soundcloud (or similar sites) can benefit from results.
NN and Clustering results can help Soundcloud fight spam.

Provided good representation for user data
Applicable to Soundcloud or other similar sites.

Open-Source Contribution
Providing Paper, Code, and Dataset for other researchers

Questions?

github.com/pmsosa/CS276-Project

github.com/netzo92/SoundCloudBotDetection