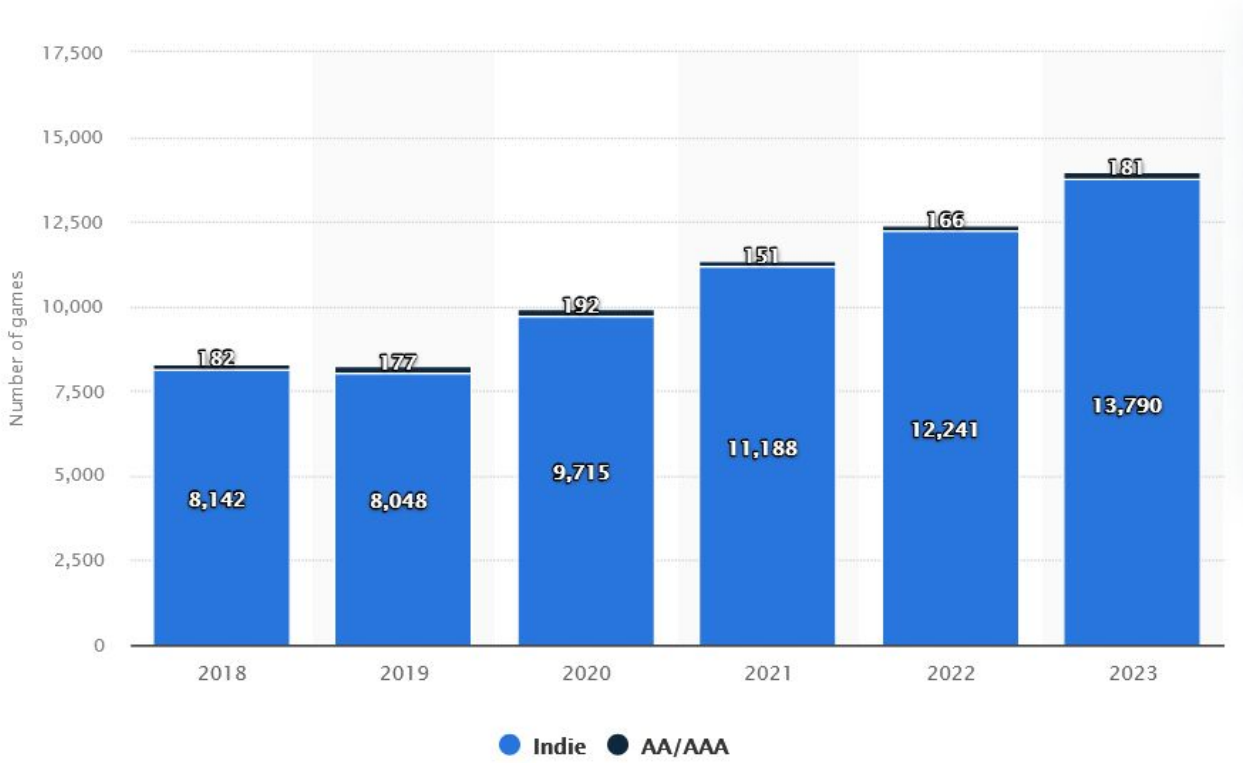
Indie Game Recommendation System

Zachary Cohen’s Capstone Project
Sprint 3

Presentation Overview

1. Intro
2. Approach
3. Exploratory Analysis
4. Results
5. Closing remarks I'll definitely have time for

Introduction



https://www.statista.com/statistics/1411839/number-games-released-steam-developer-type/

Impact approximation

132 M active users/month * 1% * $20 =

$264,000/month per percent of monthly users who purchase a game

$79,000/month (30%) to Steam

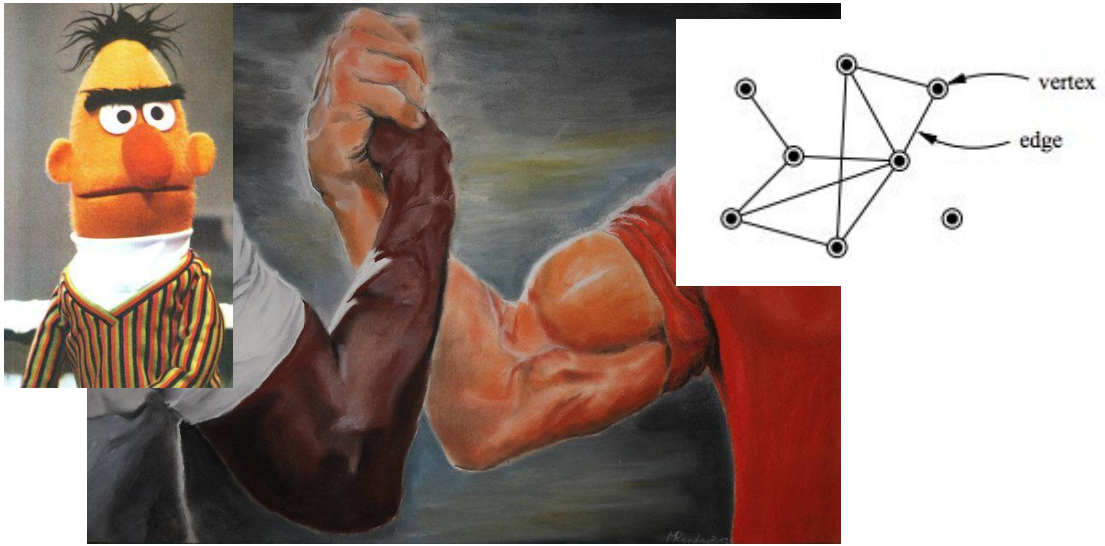
$184,800/month (70%) to indie developers

(Monthly active users from 2021 data by Statista)

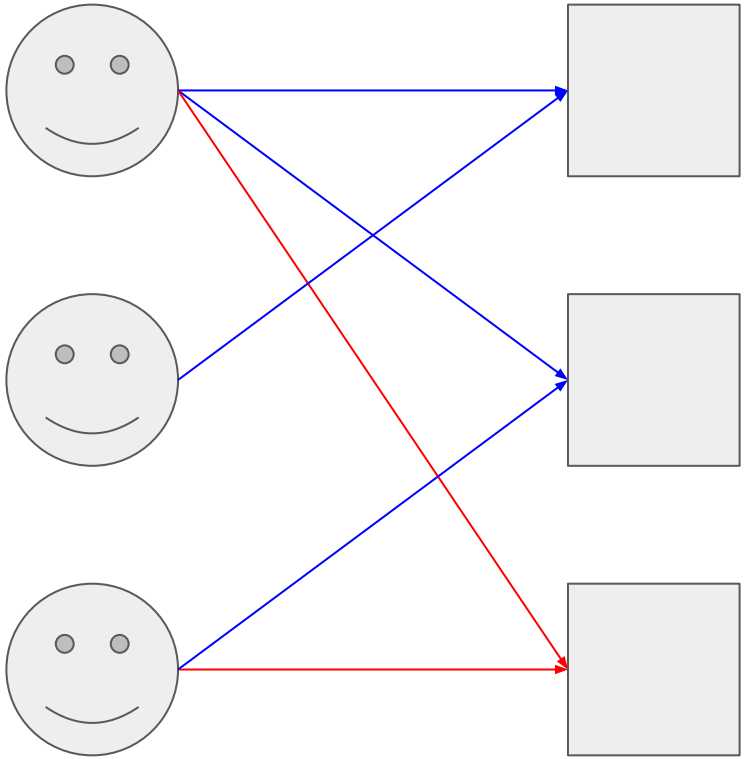
Approach

Recommender Approach

- Find user clusters (graph)
- Find game labels (BERT topic analysis)
- Recommend games via collaborative filtering of user clusters
- Filter by topic



Review Network



BERT for Tagging Games

Pros:

- Minimal Pre-processing

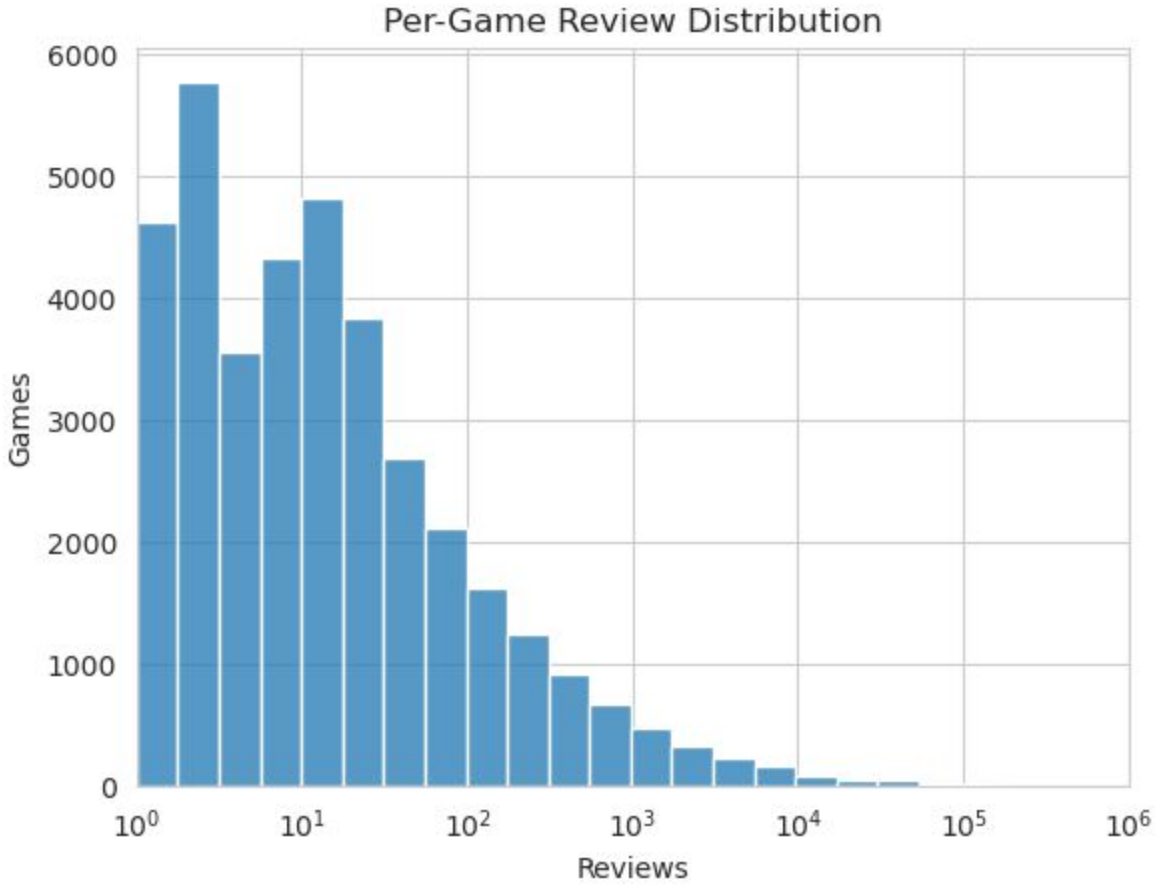
Cons:

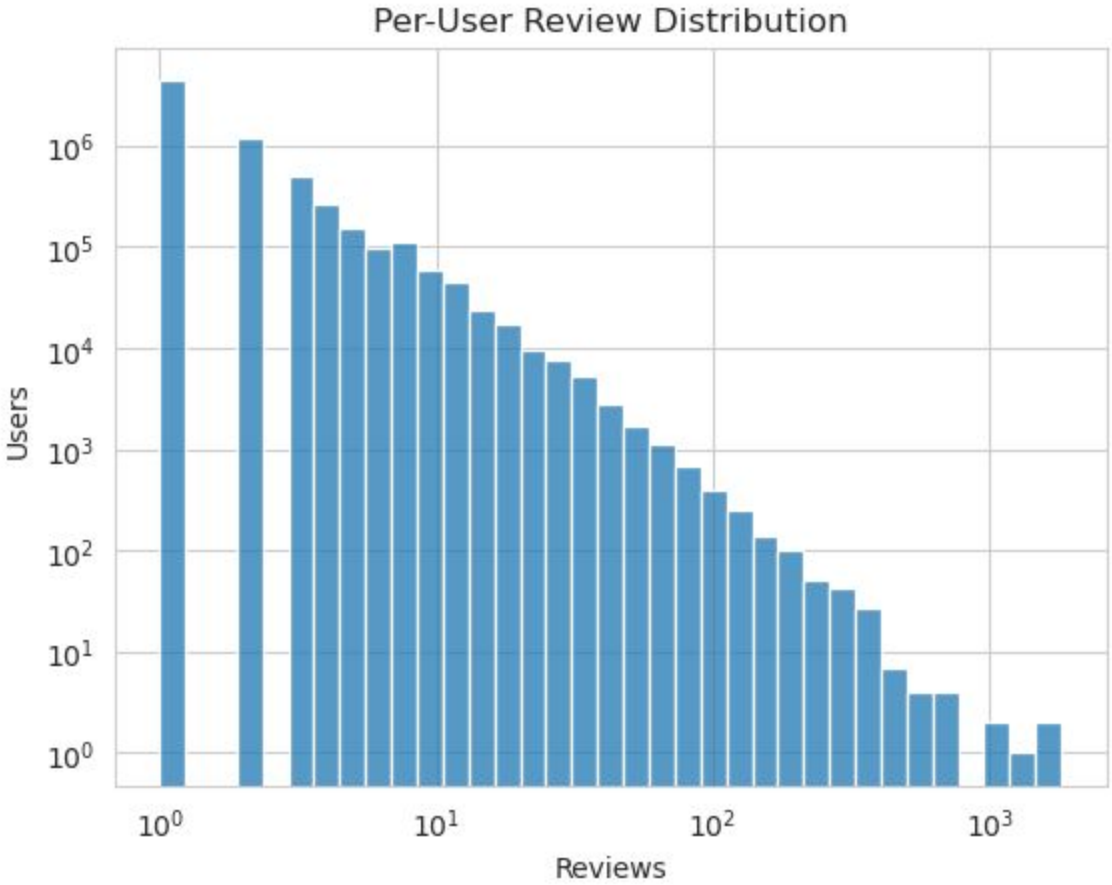
- Still needs too much memory
- Not as likeable as Ernie

Approach:

- Filter out duplicate reviews
- Use GPU acceleration (this was a pain)
- Batch-process
- Zero-shot classification

Exploratory Analysis


Per-Game Review Distribution
Games
Reviews
6000
5000
4000
3000
2000
1000
0
10^0
10^1
10^2
10^3
10^4
10^5
10^6


Per-User Review Distribution
Users
10^6
10^5
10^4
10^3
10^2
10^1
10^0
10^0
10^1
10^2
10^3
Reviews

Results

Community Clustering

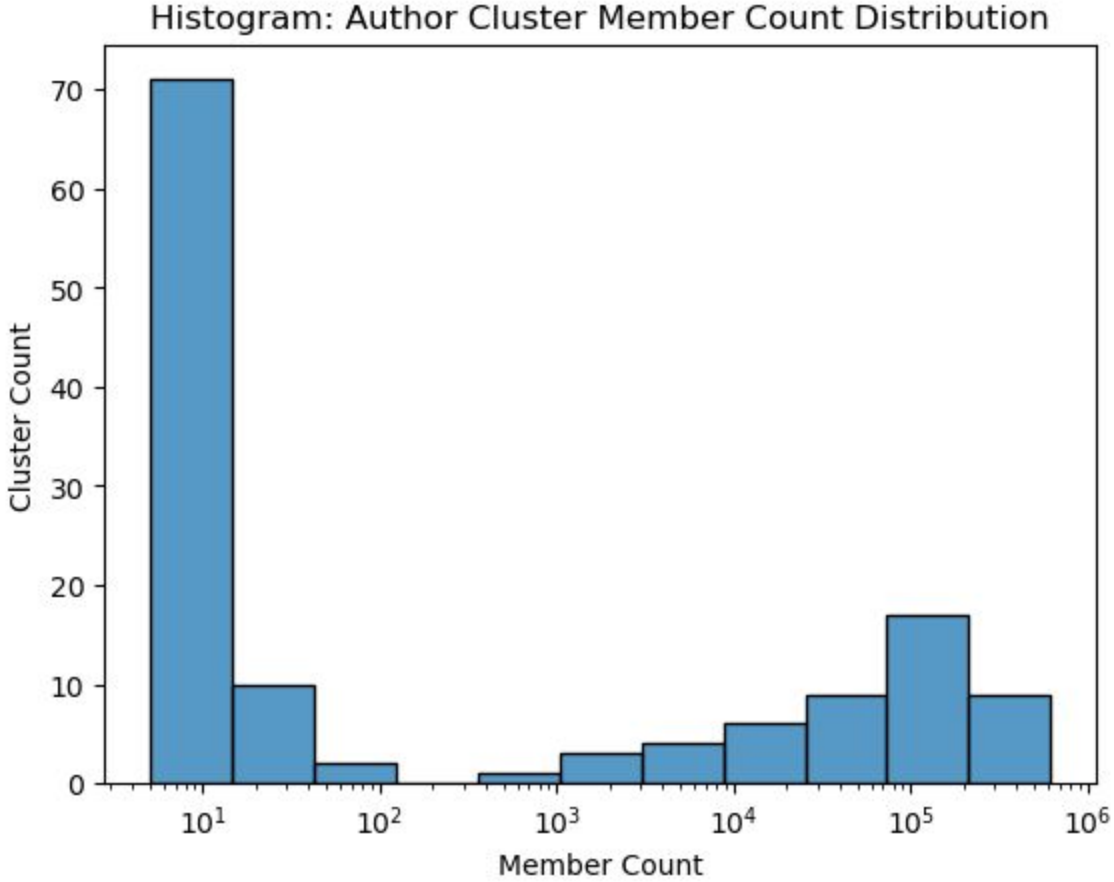
- Two bipartite graphs: positive and negative reviews
- Communities identified via Leiden algorithm (scikit-network)
  - Communities with < 5 members merged together
  - Users/games with no community all assigned to the same community

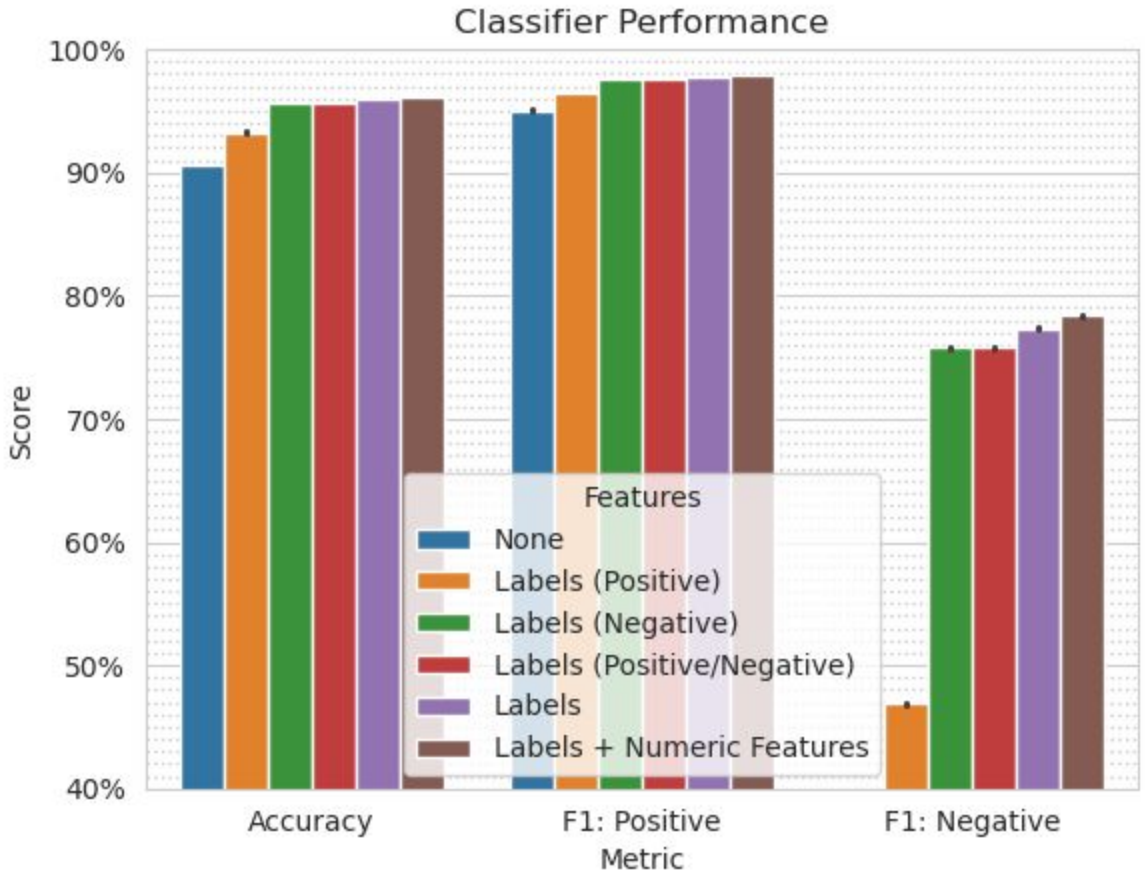
	Percent Total		Absolute		
	Positive	Negative	Positive	Negative	Total
Games	94%	74%	35,181	27,971	37,588
Authors	95%	13%	6,554,623	883,054	6,908,265
Reviews	91%	9%	13,074,002	1,357,824	14,431,826

Community Clustering

- Two bipartite graphs: positive and negative reviews
- Communities identified via Leiden algorithm (scikit-network)
  - Communities with < 5 members merged together
  - Users/games with no community all assigned to the same community

	Percent Total		Absolute		
	Positive	Negative	Positive	Negative	Total
Games	94%	74%	35,181	27,971	37,588
Authors	95%	13%	6,554,623	883,054	6,908,265
Reviews	**91%**	**9%**	13,074,002	1,357,824	14,431,826


Histogram: Author Cluster Member Count Distribution
Cluster Count
0
10
20
30
40
50
60
70
10^1
10^2
10^3
10^4
10^5
10^6
Member Count


Classifier Performance
100%
90%
80%
70%
60%
50%
40%
Score
Features
None
Labels (Positive)
Labels (Negative)
Labels (Positive/Negative)
Labels
Labels + Numeric Features
Accuracy
F1: Positive
F1: Negative
Metric

Recommender System

- Collaborative filtering of games within user cluster
  - All users if user has no cluster
- Weighs games by popularity
  - “Popularity” metric accounts for positive review percent and number of reviews

Results - Trial 1

User Reviews

app_name	voted_up
Terraria	True
Mindustry	True
FTL: Faster Than Light	True
SYNTHETIK: Legion Rising	True
Valheim	True

Recommendations

app	probability	count	voted_up	popularity	score
Noita	0.979946	18903	0.983124	0.586789	0.575022
Gunfire Reborn	0.939093	19563	0.989981	0.605849	0.568948
Hades	0.983846	16817	0.980734	0.576507	0.567194
Vampire Survivors	0.838942	60120	0.995542	0.664988	0.557886
Factorio	0.996368	9862	0.977388	0.548901	0.546908

Results - Trial 2

User Reviews

app_name	voted_up
Terraria	True
Undertale	True
Bio Inc. Redemption	True

Top 5 Recommendations

app	probability	count	voted_up	popularity	score
West of Loathing	0.974669	3686	1.000000	0.552249	0.538260
Plague Inc: Evolved	0.863236	8792	0.999204	0.594077	0.512829
Papers, Please	0.833245	13117	0.999771	0.614090	0.511688
Garry's Mod	0.915966	1735	0.976945	0.473305	0.433531
Bloons TD 6	0.964564	764	0.989529	0.446112	0.430304

Results - Trial 3

User Reviews

app_name	voted_up
Spin Rhythm XD	True

Top 5 Recommendations

	app	probability	count	voted_up	popularity	score
5535	A Hat in Time	0.758072	13624	0.995302	0.604519	0.458269
10178	Outer Wilds	0.745585	18097	0.991103	0.605677	0.451584
8288	Celeste	0.691729	24161	0.992053	0.619727	0.428683
340	Terraria	0.703419	29797	0.981944	0.599770	0.421890
8971	Dead Cells	0.701501	20396	0.987498	0.600949	0.421566

Closing Remarks

Implementation

- Ask user to name a few games they like, perhaps a game they dislike
- Ask user for genres/tags they’d be interested in
- Generate recommendations

Next Steps

- Evaluate BERT-derived tag assignments; introduce to model
  - Content-based filtering or
  - Let users select what tags they're looking for
- Tune hyperparameters
  - Explore quantitative metrics for collaborative filtering scoring
- Use negative-review derived clusters?
- Polish
  - Comments
  - Readme
  - Etc
- Package into snazzy app?

Thank you!