

# RECOMMENDER ENGINE OPTIMIZATION



**STEVEN L TRUONG**

Friday, 04/30/2021



CELEBRATING 10 YEARS OF  
**FANDOR**  
STREAMING INDEPENDENT CINEMA



CELEBRATING 10 YEARS OF  
**FANDOR**  
STREAMING INDEPENDENT CINEMA



Become an independently original program

01



CELEBRATING 10 YEARS OF  
**FANDOR**  
STREAMING INDEPENDENT CINEMA

02



Become an independently original program



Optimize the recommender engine



CELEBRATING 10 YEARS OF  
**FANDOR**  
STREAMING INDEPENDENT CINEMA

03



Become an independently original program



Optimize the recommender engine



Reduce consumers churning rate

# WHY?

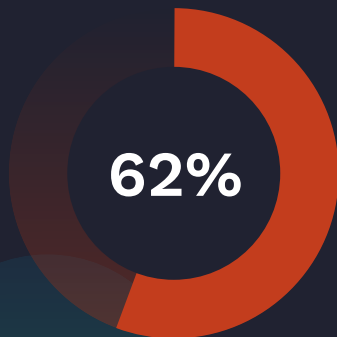
WHY DO WE NEED TO IMPROVE THE  
RECOMMENDER ENGINE?

# Motivation

According to Deloitte, among customers who cancel the services

# Motivation

According to Deloitte, among customers who cancel the services

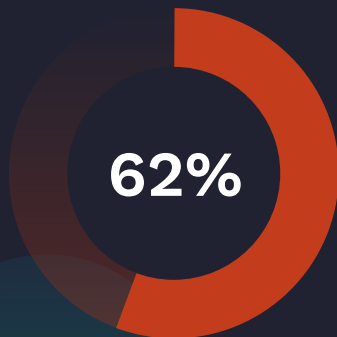


ONE SPECIFIC SHOW  
THEN CANCEL

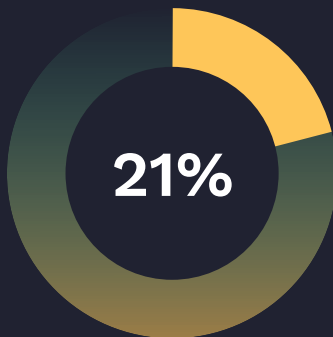


# Motivation

According to Deloitte, among customers who cancel the services



ONE SPECIFIC SHOW  
THEN CANCEL



LACKS OF NEW  
CONTENT

# Motivation

According to Deloitte, among customers who cancel the services



A donut chart with a dark blue background and a red-orange segment representing 62% of the total. The percentage '62%' is displayed in white text in the center of the donut.

62%

ONE SPECIFIC SHOW  
THEN CANCEL



A donut chart with a dark blue background and a yellow-orange segment representing 21% of the total. The percentage '21%' is displayed in white text in the center of the donut.

21%

LACKS OF NEW  
CONTENT



A donut chart with a dark blue background and a teal segment representing 27% of the total. The percentage '27%' is displayed in white text in the center of the donut.

27%

STAY IF MORE  
EXCLUSIVE CONTENT

A stylized illustration of a stage with spotlights. At the bottom, a brown stage structure with yellow trim is visible. Several spotlights shine upwards from the stage, creating a large, bright, circular area of light in the center. The word "Solution?" is written in white, bold, sans-serif font across the middle of the image. The background is dark blue with some faint, abstract shapes and a small yellow sphere on the right side.

**Solution?**

# Hypothesis and solution path

# Hypothesis and solution path

Pre-launch

Implement SOTA  
algorithms



01

A diagram illustrating the first step of a solution path. It features a large orange circle containing the number '01'. A vertical line extends upwards from the top of this circle, ending in a small orange dot. The background is dark blue with a horizontal orange band passing behind the circle.

# Hypothesis and solution path

Pre-launch

Implement SOTA  
algorithms

01



02

Pre-launch

Measure errors and A/B  
testing the engine

# Hypothesis and solution path

## Pre-launch

Implement SOTA algorithms

01



02



03

## Launch

Reduce the churning rate by 10%

## Pre-launch

Measure errors and A/B testing the engine

# Hypothesis and solution path

## Pre-launch

Implement SOTA algorithms

01



02



03



04

## Launch

Reduce the churning rate by 10%

## Pre-launch

Measure errors and A/B testing the engine

## 3-month

Reduce churning rate by 20%



# Data and tools



## Data

Acquire from  
**MovieLens** website



## EDA/Visualization

Google Sheet and  
Tableau



## Language

Python, SQL, HTML, CSS



## Packages

surprise, numpy,  
pandas, requests



## API

The Movie Database  
API



## Deployment

Flask, SQLAlchemy



HOW DO WE  
APPROACH  
THE  
PROBLEM?

# Approach

01

**User-based** collaborative filtering. Recommend similar movies based on user's preference.



# Approach

02

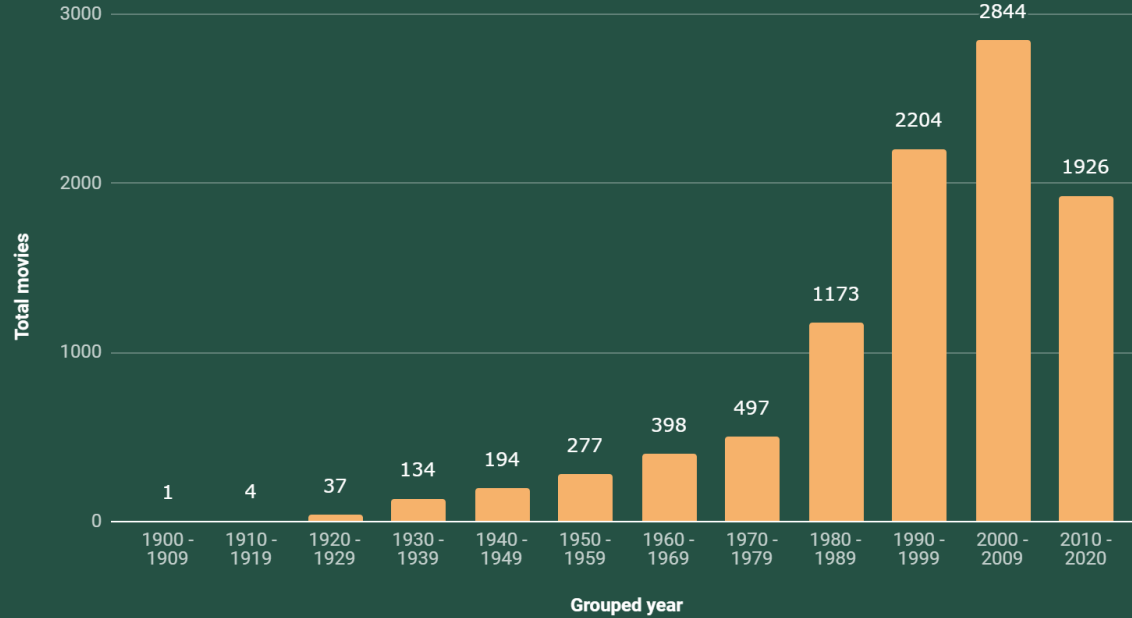
**Item-based** collaborative filtering. Recommend similar movies based on its characteristics.





# Results

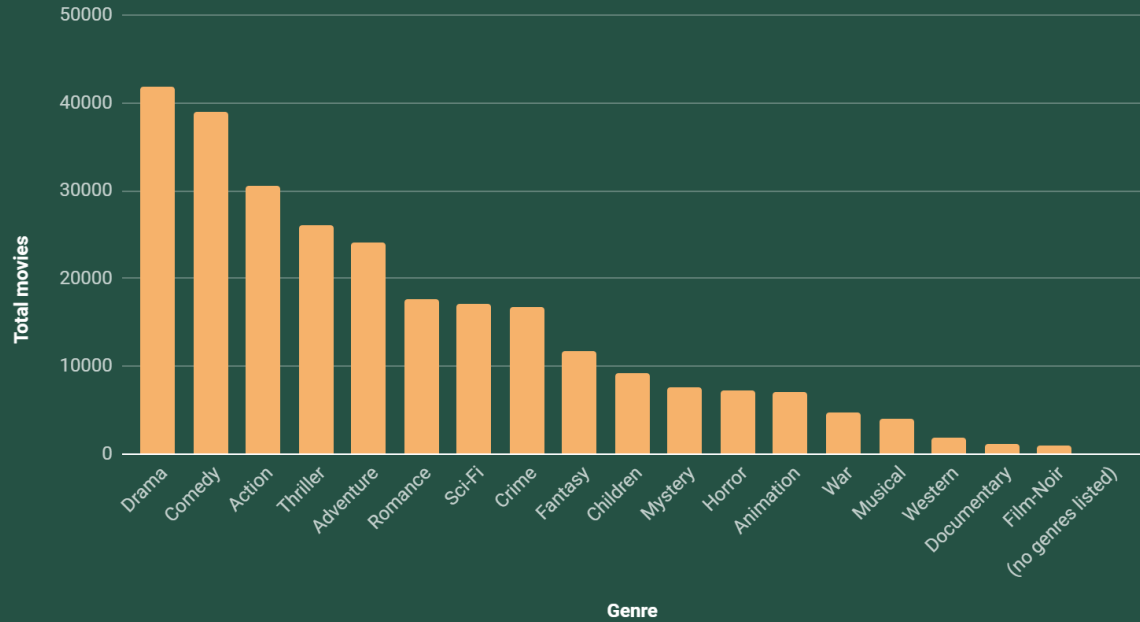
Total movies by grouped year



Total movies  
by decade



Genre vs its total movies

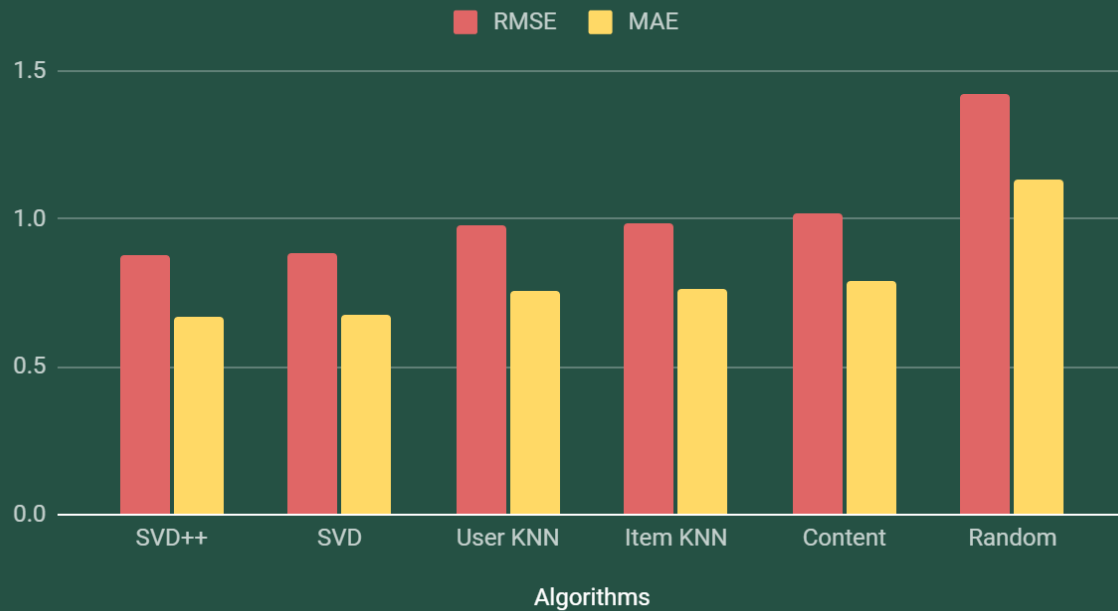


Total movies  
by genre





## RMSE and MAE Scores



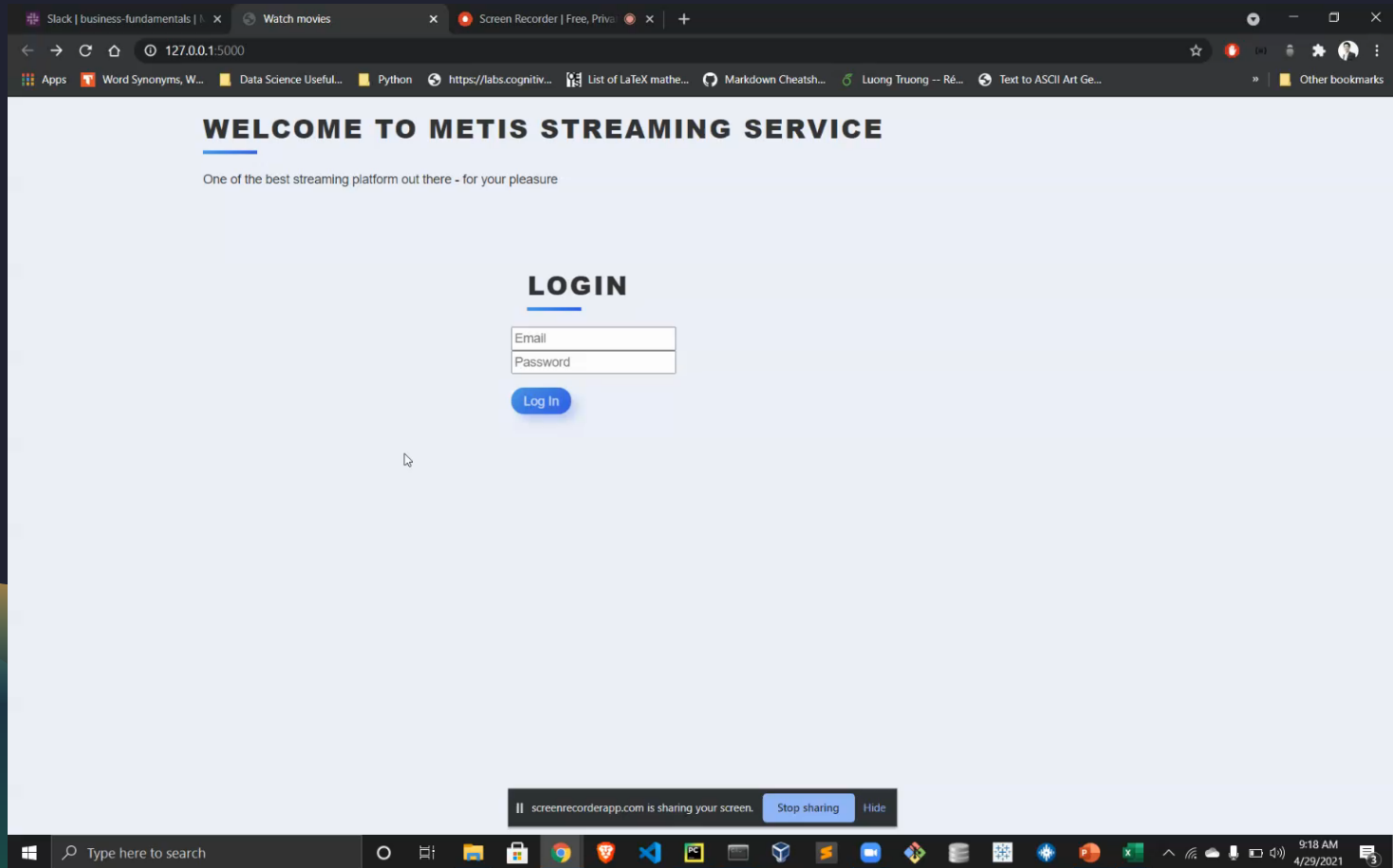
The lower the score, the better the algorithm







Demo



# Future Work

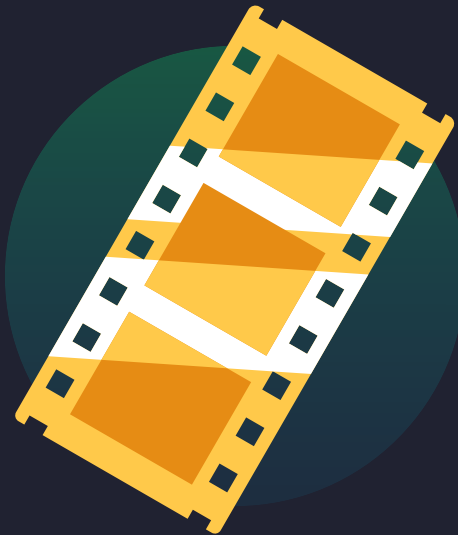
## Algorithm

Restricted Boltzmann Machines  
(RBM's)

Deep Neural Networks (RNN)

Tensorflow Recommenders (TFRS)

NLP



## Data

Scale it up using Apache  
Spark

Amazon Deep Scalable Sparse

Tensor Network Engine  
(DSSTNE)

# Thank you!

**STEVEN L TRUONG**



<https://github.com/luongtruong77>



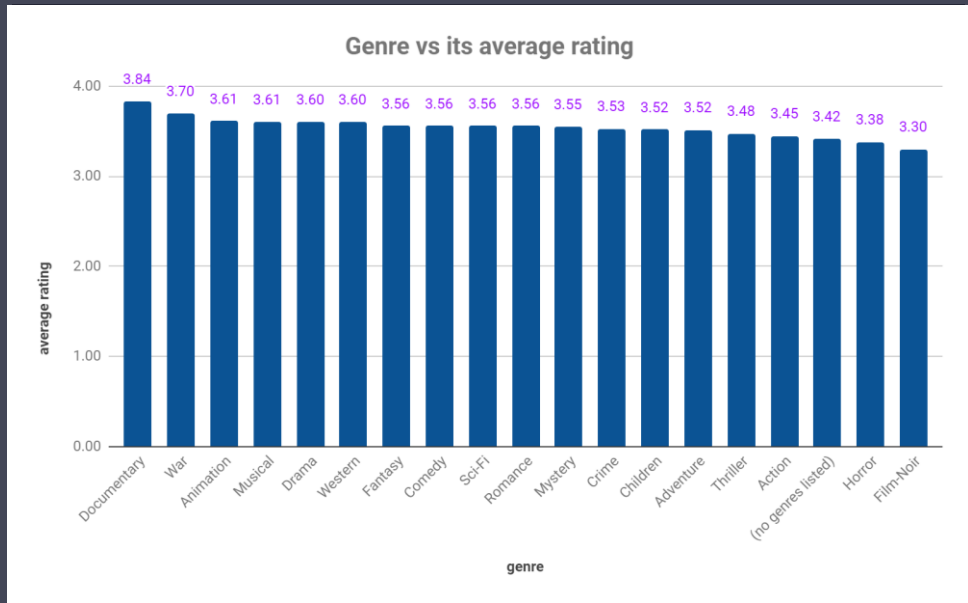
<https://www.linkedin.com/in/luongtruong77/>



# Appendix

## Average rating by genre

The Documentary genre has the highest average rating, followed by War and Animation. Horror and Film-Noir have the lowest rating of all.



# Appendix

```
1 df.sample(20)
```

executed in 31ms, finished 12:18:42 2021-04-29

	movieId		title	year	userId	rating	sub_genre_1	sub_genre_2	sub_genre_3	sub_genre_4	sub_genre_5
68958	5943		Maid in Manhattan (2002)	2002	606	3.0	Comedy	Romance	None	None	None
90223	71902		Spread (2009)	2009	432	3.5	Drama	Romance	None	None	None
8924	318		Shawshank Redemption, The (1994)	1994	517	4.0	Crime	Drama	None	None	None
8093	296		Pulp Fiction (1994)	1994	437	5.0	Comedy	Crime	Drama	Thriller	None
23018	1080		Monty Python's Life of Brian (1979)	1979	608	2.0	Comedy	None	None	None	None
55863	3618		Small Time Crooks (2000)	2000	448	4.0	Comedy	Crime	None	None	None
15567	588		Aladdin (1992)	1992	330	3.0	Adventure	Animation	Children	Comedy	Musical
32496	1527		Fifth Element, The (1997)	1997	477	4.0	Action	Adventure	Comedy	Sci-Fi	None
27579	1246		Dead Poets Society (1989)	1989	318	3.5	Drama	None	None	None	None
23457	1094		Crying Game, The (1992)	1992	95	4.0	Drama	Romance	Thriller	None	None
71247	6523		Mr. Baseball (1992)	1992	599	2.0	Comedy	Romance	None	None	None
12747	457		Fugitive, The (1993)	1993	470	4.0	Thriller	None	None	None	None
6973	260		Star Wars: Episode IV - A New Hope (1977)	1977	292	4.0	Action	Adventure	Sci-Fi	None	None
85098	53322		Ocean's Thirteen (2007)	2007	28	1.5	Crime	Thriller	None	None	None
19276	788		Nutty Professor, The (1996)	1996	151	5.0	Comedy	Fantasy	Romance	Sci-Fi	None
45896	2628		Star Wars: Episode I - The Phantom Menace (1999)	1999	382	3.5	Action	Adventure	Sci-Fi	None	None
59948	4036		Shadow of the Vampire (2000)	2000	414	3.0	Drama	Horror	None	None	None
99862	158238		The Nice Guys (2016)	2016	249	4.5	Crime	Mystery	Thriller	None	None
43136	2396		Shakespeare in Love (1998)	1998	201	5.0	Comedy	Drama	Romance	None	None
95227	99007		Warm Bodies (2013)	2013	125	4.5	Comedy	Horror	Romance	None	None

**20 random rows from  
over 100,000 rows**

Each movie has one or more genres associated with it. For example, Aladdin (1992) is considered “Adventure | Animation | Children | Comedy | Musical” movie whereas Dead Poets Society (1989) is just a Drama.

# Appendix

$$\text{Similarity}(p, q) = \cos \theta = \frac{p \cdot q}{\|p\| \|q\|} = \frac{\sum_{i=1}^n p_i q_i}{\sqrt{\sum_{i=1}^n p_i^2} \sqrt{\sum_{i=1}^n q_i^2}}$$

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}}$$

**Cosine similarity (top) and adjusted cosine similarity (bottom) formulas**

We use cosine similarity to measure the similarity between movies (they are identical if  $\alpha$  is 0 degree and totally different if  $\alpha$  is 90 degree).