Insert the sub title of your presentation

Non-Personalised Video Games Recommender

Kenneth Lim / DSI 35

Contents

Background

Recommender Systems

Online Information retrieval systems aimed at providing personalised recommendations to users based on preferences or interests

Widely integrated in many applications and platforms to keep users engaged

Aims to provide a concise summary of products for users to choose from

Applications of Recommender Systems

Ecommerce

Amazon

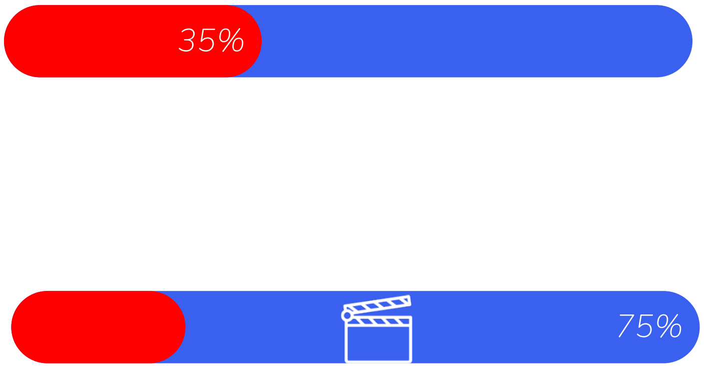
Movie Streaming

Netflix

Dating Apps

Tinder

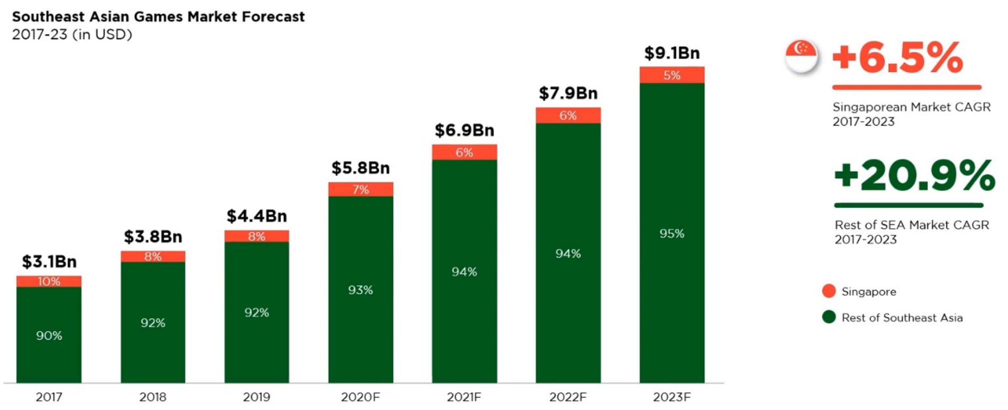
Consumer Benefits of Recommender Systems



35% of consumer purchases on Amazon
75% of show titles watched on Netflix
Generate new business revenue through personalisation

[McKinsey]

Gaming Industry



Singapore's Esports sector

1. Rising industry due to huge market growth during Covid-19 pandemic

2. Projected to have an compounded annual revenue growth of about 6.5%

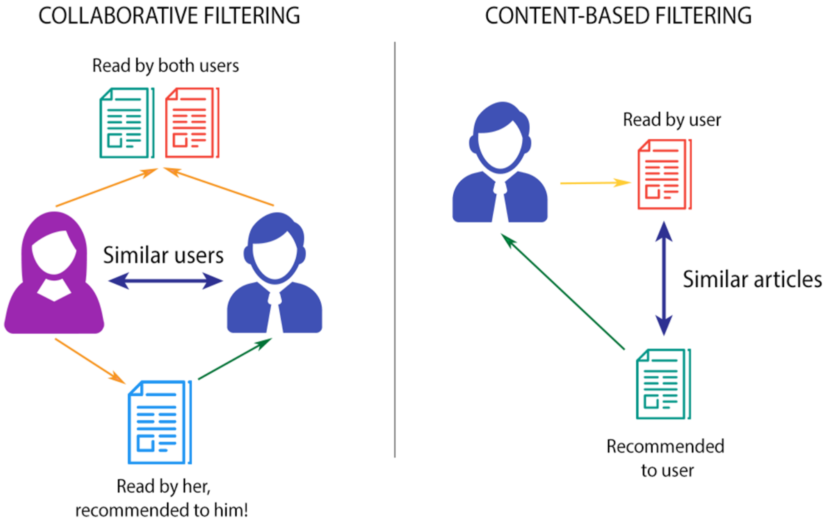
3. Global average weekly playtime on video games: 8-9 hours

Singapore's gaming industry

Techrepublic

Objective: Build non-personalised video games recommender

Mechanism of Recommender Systems



Content filtering: Recommend similar titles to users (Item-Item)

Collaborative filtering: Recommend similar titles based on user's preferences and behaviour (User-User)

EDA

Dataset

Games

46K

46000 unique game titles with game information till Feb 2023

Metadata

46K

46000 rows of unique description with game genre tags Feb 2023

Recommendations

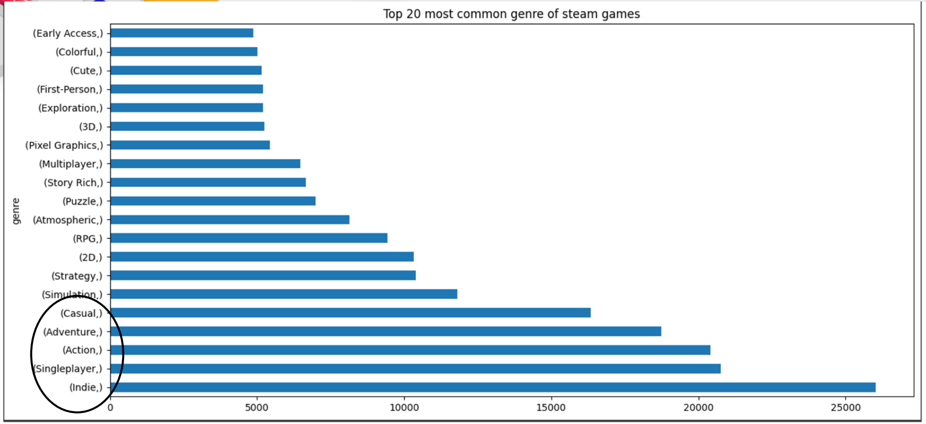
10M

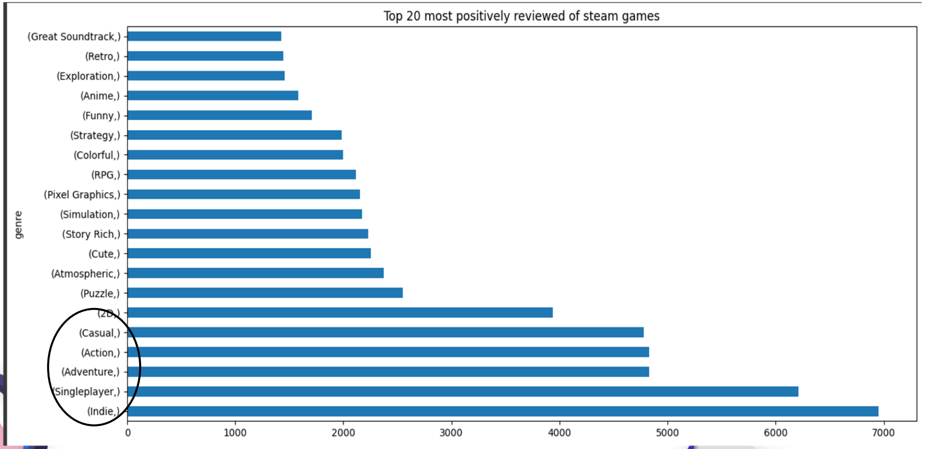
10 million reviews from 5.8 million unique users Dec 2022

Final dataset

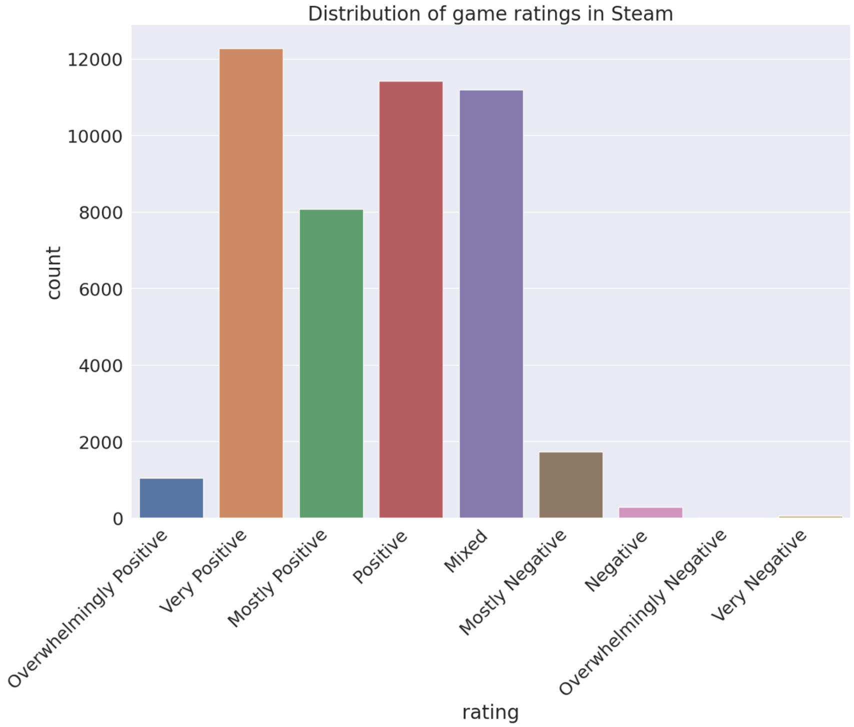
~700k

700k reviews from 40k unique users

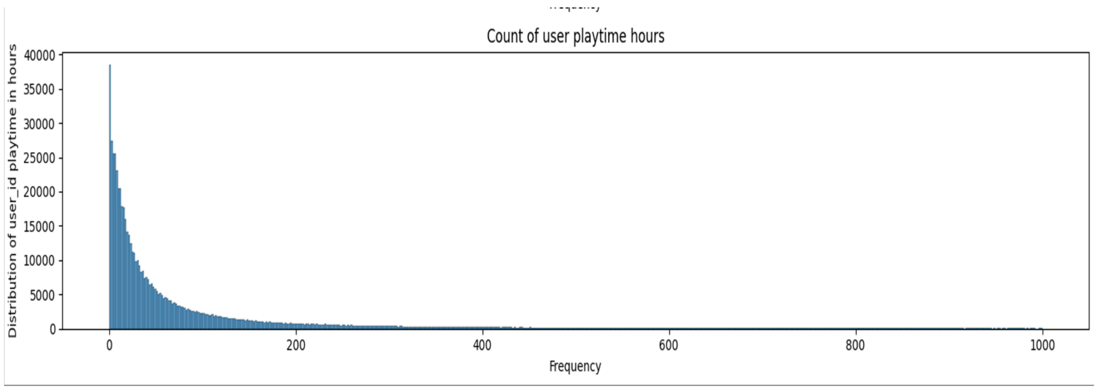




Action, Adventure, SinglePlayer, Indie and Casual genres most popular



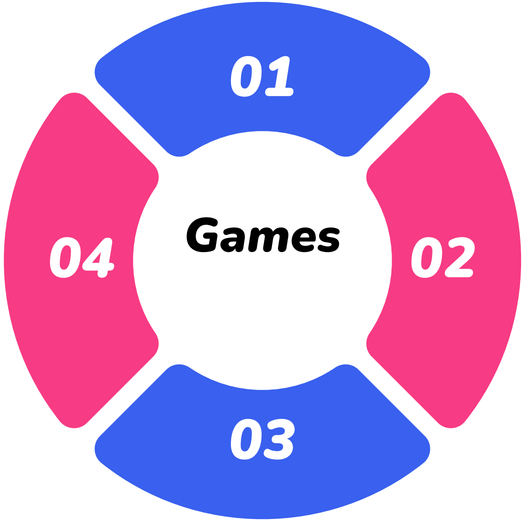
Most games are rated to be positive on overall



Most users have an overall average playtime of less than 100 hours

Data Pre-processing

Content Filtering



Merge games and metadata dataset

Extract keywords from game description

Using RAKE (Rapid Automatic Keyword Extraction) library in NLTK python

Combine textual information

Large text of word corpus for NLP modelling

Cosine Similarity

Count Vectorizer and TF-IDF Vectorizer

Collaborative Filtering

Binning transformation + rank scaling

- *Group the users' playtime into 10 equal bins*
- *Scale the bin groups into range of 1 to 5*
- *Transformed feature will be directly inferred as scale of continuous ratings from users' reviews*

Playtime inference for explicit ratings

- *Steam does not provide platform for numerical ratings*
- *Quantify relevance of each game title to individual users*
- ***If the user play a game for a long time: he/she likes the game***

Modelling and Evaluation

Content Filtering

Vectorizer: TF-IDF, ngram = (1,1)

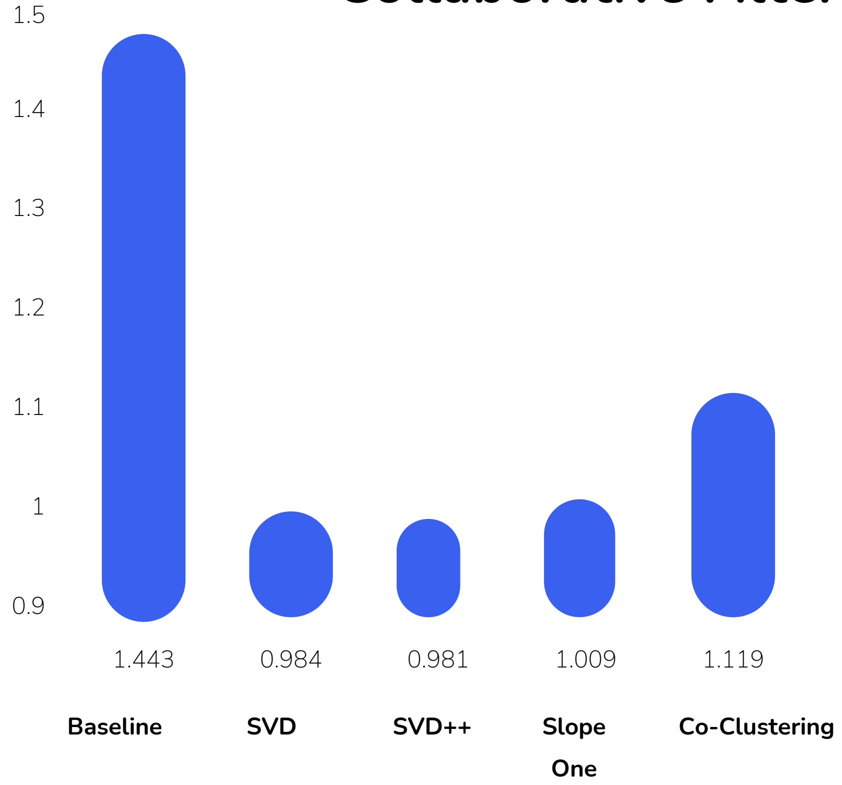
Assassin's Creed® Odyssey

index	title	similarity_score
0	Assassin's Creed® III Remastered	0.314
1	Assassin's Creed® Origins	0.296
2	Discovery Tour by Assassin's Creed®: Ancient Egypt	0.283
3	Assassin's Creed® Liberation HD	0.277
4	Assassin's Creed® Chronicles: China	0.263
5	Assassin's Creed® Unity	0.262
6	Assassin's Creed™: Director's Cut Edition	0.249
7	Assassin's Creed® Rogue	0.24
8	Notre-Dame de Paris: Journey Back in Time	0.22
9	Prince of Persia®	0.198

Sid Meier's Civilization® III Complete

index	title	similarity_score
0	Civilization IV: Beyond the Sword	0.341
1	Sid Meier's Covert Action (Classic)	0.322
2	Silent Service	0.322
3	Sid Meier's Civilization IV: Colonization	0.315
4	Civilization IV®: Warlords	0.291
5	Sid Meier's Civilization® IV	0.253
6	Hero Generations	0.234
7	Sid Meier's Railroads!	0.231
8	Sid Meier's Civilization®: Beyond Earth™	0.215
9	Don't Kill the Cow	0.154

Collaborative Filtering (RMSE)



SVD uses matrix factorization to decompose user-item matrix into individual user and item matrices

Identify latent features influencing interactions between users and products

Collaborative Filtering (Precision@10, Recall@10)

Model No.	Model	Precision@10	Recall@10
1	Random Predictor (Baseline)	0.454	0.505
2	**SVD (Matrix Factorization)**	0.557	0.734
3	SVD++ (Matrix Factorization)	0.558	0.735
4	SlopeOne	0.537	0.691
5	Co-Clustering	0.435	0.443

Goal: provide comprehensive list of relevant game titles

Optimise Recall@10 meets the project requirements

$$\text{Precision@k} = \frac{|\{\text{Recommended items that are relevant}\}|}{|\{\text{Recommended items}\}|}$$

$$\text{Recall@k} = \frac{|\{\text{Recommended items that are relevant}\}|}{|\{\text{Relevant items}\}|}$$

Collaborative Filtering

Model: SVD

Hyperparameters: n_factors = 20, n_epochs =10, lr_all = 0.005, reg_all=0.05

User ID: 4924848

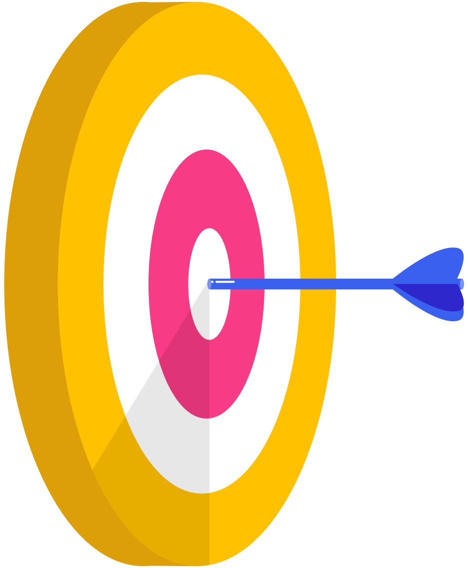
index	title
0	Panzer Corps 2
1	Total War: WARHAMMER
2	Total War: WARHAMMER III
3	FINAL FANTASY XIV Online
4	Monster Hunter: World
5	Deep Rock Galactic
6	Yakuza: Like a Dragon
7	Cyberpunk 2077
8	MONSTER HUNTER RISE
9	Pathfinder: Wrath of the Righteous - Enhanced Edition

User ID: 3869605

index	title
0	Zero Escape: The Nonary Games
1	Danganronpa V3: Killing Harmony
2	Sid Meier's Civilization® V
3	Terraria
4	Total War: WARHAMMER
5	Crusader Kings III
6	Kingdom Come: Deliverance
7	Monster Hunter: World
8	MONSTER HUNTER RISE
9	Pathfinder: Wrath of the Righteous - Enhanced Edition

Conclusion, Limitations and Future Work

Conclusions



Goal 01.

Built a prototype NLP Content Filtering model to accurately recommend game titles to users based on their past activity

Goal 02.

Built a prototype Collaborative Filtering model to accurately recommend game titles based on users' activity

Advantages

Content Filtering: User-focused recommendations, Scalable to wide user pool individually

Collaborative Filtering: Takes into account user with similar interest and preferences to recommend titles to new users

Disadvantages

Content Filtering: Cold Start, User specific centric

Collaborative Filtering: Sparsity, Scalability

Future Work

Deploy model on Cloud platform

Scaling up computing resources to address memory limitations on large datasets

Employ Implicit prediction algorithms

More accurate predictions of users' preferences from implicit features

Employ hybrid based recommenders

Address cold start issue and matrix sparsity

Thank You

Streamlit demo