STEAM GAMES RECOMMENDER SYSTEMS

A Project Report

Submitted in Partial Fulfillment of Requirements for Sharpest Minds Mentorship Program

Submitted By – Swaroop Todankar

Date – 18th Dec 2019

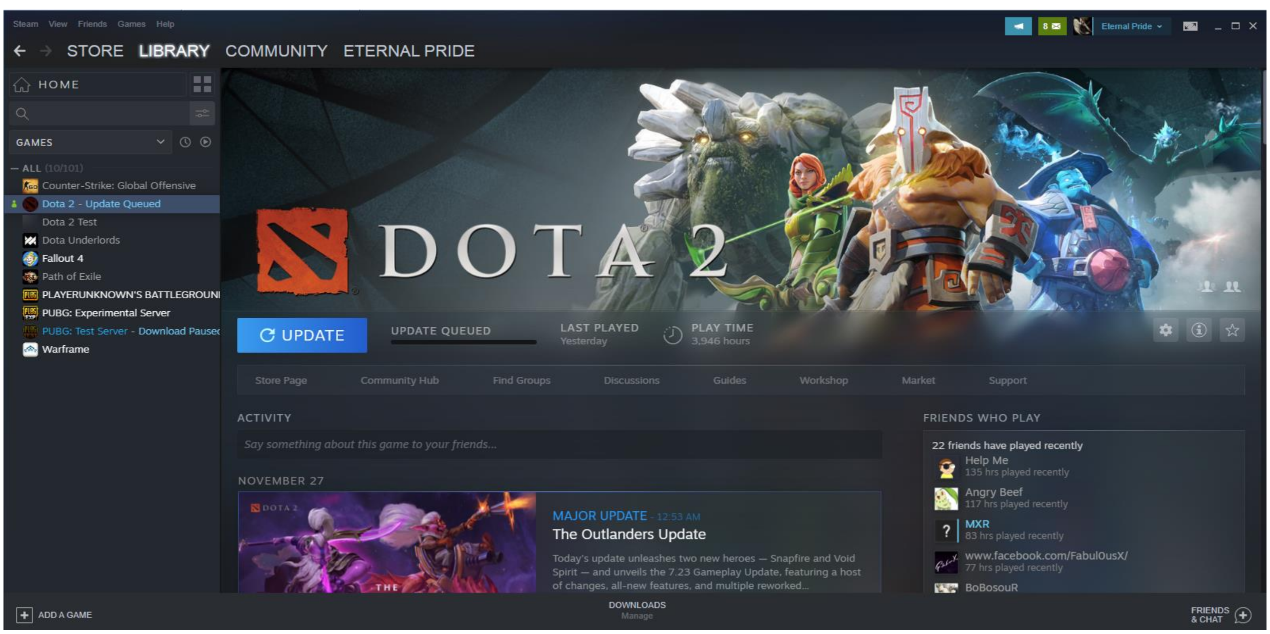

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction

Steam is a video game digital distribution service provided by Valve. It is the largest platform in the video game distribution genre by player base. [1]

Users of steam can download games digitally in to their profile library. Games can be purchased and gifted between users and these games stay until deleted by the user himself.

The Kaggle completion involving dataset for steam users and the games in their libraries was used for the analysis. Further information provided was the amount of hours a particular user has played the game. The aim of the analysis was to build a recommender system which could accurately suggest games to other users which they haven't played yet.

2. Business Problem

The main question is- depending upon the number of games purchased by the users and the amount of time in hours spent in playing the games, what different games can be suggested to different users using a recommender system?

This report aims to put forward an analysis of building a recommender system to recommend games to users with similar interests.

3. People interested in the Project – Target Audience

The target audience in this scenario are the people – gamers, who would like to be recommended with similar games according to their interests.

4. Data required

The data required to build a model to predict the customer satisfaction is as follows:

User Data: Data pertaining to the users and games played is provided by the Kaggle user for analysis.

5. Methodology

The following steps were employed to obtain the required results:

5.1 Importing Necessary Libraries

The first step is to import the necessary libraries and packages.

- Numpy – For numerical calculations
- Pandas – Data manipulation
- Turicreate – Machine Learning library for Recommender Systems

5.2 Google Drive Pre- requisites

This step involves the authentication steps taken in order to use the dataset stored in Google Drive which can be directly loaded into Google Colab.

5.3 Pre-Processing

The following preprocessing steps were employed:

1. Checking the shape of the dataset

2. Checking the data types of the columns (features) present in the dataset

3. Removal of unwanted columns

4. Getting statistical information of the dataset

5. Displaying brief information of the dataset

6. Creating a new dataframe as per requirement by merging

5.4 Ranking

The dataset provided is without a rating column for the games played by the user. The next step included creation of a ranking system depending upon the game hours played by a user.

This ranking system will be used as a rating system by grouping the dataset by users and games and then using the (pd.cut) command on it. This results in a rating of 10 for the top games played by the user and the remaining accordingly.

5.5 Recommender Systems – User Recommendations

Using the Turicreate library provided by Apple, the following three recommender systems were built:

1. Factorization Recommender
2. Ranking Factorization Recommender
3. Item Similarity Recommender

5.5.1 Creation of SFrame

The use of Turicreate library requires a SFrame which is created from the original dataframe.

5.5.2 Building the Recommender Systems

<u>1. Factorization Recommender –</u> Observed ratings are modeled as a weighted combination of terms, where weights are learned from data.

<u>2. Ranking Factorization Recommender</u> – Recommends items that are both similar to items in a user's dataset and if rating is given, which item would be rated highly by the user.

<u>3. Item Similarity Recommender</u> – Computes the similarity between each pair of items and recommends items to each user that are closest to items they have already used and liked. [2]

5.5.3 Display the results

The recommendations for the top 5 users are obtained and compared in the Result section.

5.6 Reduction of Dataset

The dataset is optimized by removal of all the users with less than 5 games in their steam library.

5.7 Recommendation Systems – Testing and Evaluation

This step involves the following steps:

5.7.1 Convert into SFrame

The optimized dataset is again converted in to a SFrame.

5.7.2 Splitting the optimized dataset into Train and Test set

Obtained SFrame is split into a Training set and a Test set.

5.7.3 Training and Evaluation

The model is trained on the training set for all the three recommender systems and then model performance is evaluated on the testing set using RMSE as a metric.

5.7.4 Comparison of improvement in model performance

The results obtained from this step are presented in the result section where the performance of the models are compared.

5.8 Metrics for Evaluation

The following metrics were used for evaluation:

1. RMSE

6. Results and Discussion

6.1 Outputs/ Recommendations of 5.5 – Recommendation System – User Recommendations

The results of Approach 1 are discussed below. It includes the outputs of all 3 recommendation systems clubbed together for one single user. The results are arranged in ascending order with the top recommendation at the top with the maximum score and then the following afterwards. The column headings consists of acronyms which denote the following:

FR – Factorization Recommender

RFR – Ranking Factorization Recommender

ISR – Item Similarity Recommender

6.1.1 – Top user 1

	Game_FR	score_FR	Game_RFR	score_RFR	Game_ISR	score_ISR
0	Sid Meier's Civilization V	4.357136	Counter-Strike Global Offensive	3.167137	Borderlands 2	0.088440
1	Counter-Strike Global Offensive	4.046952	Counter-Strike	3.095890	The Witcher 2 Assassins of Kings Enhanced Edition	0.086573
2	Call of Duty Modern Warfare 2 - Multiplayer	3.700310	Counter-Strike Source	3.059422	Portal 2	0.085423
3	Grand Theft Auto V	3.682324	Unturned	2.993581	Metro 2033	0.083152
4	Football Manager 2012	3.643091	Sid Meier's Civilization V	2.371592	Dishonored	0.081220
5	Half-Life 2 Lost Coast	3.568467	Terraria	2.125734	Batman Arkham City GOTY	0.079579
6	Sniper Ghost Warrior	3.507680	Call of Duty Modern Warfare 2 - Multiplayer	2.026284	Company of Heroes (New Steam Version)	0.074126
7	Counter-Strike	3.463431	Call of Duty Modern Warfare 2	2.019285	Company of Heroes	0.068913
8	Euro Truck Simulator 2	3.399293	Grand Theft Auto V	1.969556	Batman Arkham Asylum GOTY Edition	0.067708
9	Counter-Strike Source	3.368751	Call of Duty Black Ops	1.969426	Deus Ex Human Revolution	0.067490

Figure 2: Top user 1 Recommendations

6.1.2 Top user 3

TOP 2 USER TOTAL COMPARISON FOR RECOMMENDATIONS

```
[ ] Top2_comparison = pd.concat([fact2, ranfact2,itsim2], axis=1, sort=False)
    Top2_comparison.head(10)
```

	Game_FR	score_FR	Game_RFR	score_RFR	Game_ISR	score_ISR
0	UberStrike	3.586367	Unturned	1.841668	Brtal Legend	0.021146
1	Football Manager 2012	3.378578	Grand Theft Auto V	1.195544	Overlord Raising Hell	0.021017
2	Unturned	3.345255	Call of Duty Modern Warfare 2 - Multiplayer	1.056746	Mark of the Ninja	0.020263
3	Serious Sam HD The Second Encounter	3.244238	Call of Duty Modern Warfare 2	0.959383	Bastion	0.019053
4	Call of Duty Black Ops	3.207467	Call of Duty Black Ops	0.948379	Darksiders	0.018423
5	BLOCKADE 3D	3.117440	Fallout New Vegas	0.903869	Dishonored	0.015959
6	Football Manager 2013	3.060799	PAYDAY 2	0.896814	LIMBO	0.014379
7	Call of Duty Modern Warfare 2	2.902094	Fallout 4	0.874074	FTL Faster Than Light	0.014343
8	Lost Planet Extreme Condition	2.856654	Rust	0.815132	Rise of the Argonauts	0.013883
9	F1 2012	2.819399	Dark Souls Prepare to Die Edition	0.751710	The Walking Dead	0.013651

Figure 3: Top user 2 Recommendations

6.1.3 Top user 3

TOP 3 USER TOTAL COMPARISON FOR RECOMMENDATIONS

```
[ ] Top3_comparison = pd.concat([fact3, ranfact3,itsim3], axis=1, sort=False)
    Top3_comparison.head(10)
```

	Game_FR	score_FR	Game_RFR	score_RFR	Game_ISR	score_ISR
0	Sid Meier's Civilization V	4.441703	Counter-Strike Global Offensive	3.197938	GoD Factory Wingmen	0.021197
1	Counter-Strike Global Offensive	4.134893	Counter-Strike Source	2.678136	Guacamelee! Gold Edition	0.020847
2	Counter-Strike Source	3.645814	Unturned	2.387506	The Swapper	0.019165
3	Football Manager 2012	3.602747	Sid Meier's Civilization V	1.868469	Titan Quest	0.018600
4	Counter-Strike Nexon Zombies	3.501396	Call of Duty Modern Warfare 2 - Multiplayer	1.758601	Gray Matter	0.017633
5	UberStrike	3.433347	Grand Theft Auto V	1.647473	Orcs Must Die! 2	0.017199
6	Sniper Ghost Warrior	3.432557	Borderlands 2	1.519177	The Witcher 2 Assassins of Kings Enhanced Edition	0.016779
7	Call of Duty Modern Warfare 2 - Multiplayer	3.373797	Call of Duty Black Ops	1.470338	The Walking Dead	0.016545
8	Call of Duty Black Ops	3.236289	PAYDAY 2	1.456703	Overlord Raising Hell	0.015891
9	Football Manager 2013	3.235437	The Witcher 2 Assassins of Kings Enhanced Edition	1.428424	Shank 2	0.015699

Figure 4: Top user 3 Recommendations

6.1.4 Top user 4

▾ TOP 4 USER TOTAL COMPARISON FOR RECOMMENDATIONS

```
Top4_comparison = pd.concat([fact4, ranfact4,itsim4], axis=1, sort=False)
Top4_comparison.head(10)
```

	Game_FR	score_FR	Game_RFR	score_RFR	Game_ISR	score_ISR
0	Counter-Strike Global Offensive	5.088228	Dota 2	4.757342	Sine Mora	0.023019
1	Dota 2	5.080670	Counter-Strike Global Offensive	4.111945	Dishonored	0.021352
2	Grand Theft Auto V	4.513086	Team Fortress 2	4.089925	Dear Esther	0.019962
3	Unturned	4.390761	Counter-Strike	3.472717	BioShock Infinite	0.019124
4	Call of Duty Modern Warfare 2 - Multiplayer	4.301606	Counter-Strike Source	3.420557	Gratuitous Space Battles	0.017178
5	Football Manager 2012	4.229529	Garry's Mod	3.069335	Osmos	0.016514
6	Counter-Strike Source	4.173589	Unturned	2.921819	Intrusion 2	0.016453
7	Counter-Strike Nexon Zombies	4.131648	Call of Duty Modern Warfare 2 - Multiplayer	2.642764	Red Faction Guerrilla Steam Edition	0.015399
8	Call of Duty Black Ops	4.018089	Grand Theft Auto V	2.481494	Proteus	0.015194
9	F1 2012	3.990187	Empire Total War	2.285131	Thomas Was Alone	0.014901

Figure 5: Top user 4 Recommendations

6.1.5 Top user 5

▾ TOP 5 USER TOTAL COMPARISON FOR RECOMMENDATIONS

```
Top5_comparison = pd.concat([fact5, ranfact5,itsim5], axis=1, sort=False)
Top5_comparison.head(10)
```

	Game_FR	score_FR	Game_RFR	score_RFR	Game_ISR	score_ISR
0	Grand Theft Auto V	4.697976	Counter-Strike	3.885623	Combat Arms	0.025561
1	Football Manager 2012	4.267308	Grand Theft Auto V	2.796779	Dragon Nest Europe	0.022592
2	Euro Truck Simulator 2	4.247729	Empire Total War	2.777390	Global Agenda	0.020098
3	F1 2012	4.076017	Euro Truck Simulator 2	2.756396	Mark of the Ninja	0.017079
4	H1Z1	4.069415	Total War SHOGUN 2	2.620009	Batman Arkham City GOTY	0.016512
5	Sniper Ghost Warrior	4.047849	Call of Duty Black Ops II - Multiplayer	2.609438	Batman Arkham Asylum GOTY Edition	0.016301
6	Half-Life Blue Shift	4.043885	Counter-Strike Condition Zero	2.592512	Frozen Synapse	0.016037
7	Counter-Strike	3.989727	PAYDAY 2	2.589665	Dear Esther	0.015710
8	RACE 07	3.884441	Arma 2 Operation Arrowhead	2.578763	FTL Faster Than Light	0.015592
9	Team Fortress Classic	3.876007	Call of Duty Modern Warfare 3 - Multiplayer	2.548212	Dishonored	0.015285

Figure 6: Top user 5 Recommendations

6.2 Outputs/ Recommendations of 5.7 – Recommendation System – Testing and Evaluation

In this section, the output of the section 5.7 is presented. Train and test split is performed on the original and reduced datasets for all three recommender systems and the RMSE values are compared.

	Factorization Recommender RMSE	Ranking Factorization Recommender RMSE	Item Similarity RMSE
Actions			
Original	4.499	8.479	3.004
Cleaned	4.128	5.526	2.414

Figure 7: Recommender Systems Comparison

With the removal of users with very less amount of games played, the RMSE decreases considerably.

For Factorization Recommender the percentage decrease is 8.2 percent.

For Ranking Factorization Recommender the percentage decrease is 34.82 percent.

For Item Similarity Recommender the percentage decrease is 19.64 percent.

Thus the performance of Ranking Factorization Recommender increases considerably by the removal of redundant information.

7. Limitations

The model performances could have been further improved if the actual ratings provided by users were available.

8. Conclusion

Using Google Collab, the dataset stored in google drive was loaded. Pre-processing steps were carried out and using Turicreate library, 3 recommender systems were built. Dividing the data into training set and testing set, the performance of these systems were compared.

The dataset was further optimized and the process was repeated. The evaluated model performances were compared for unoptimized and optimized models. RMSE score improves with the removal of users with less than 5 games played.

9. References

1. Steam, “Steam (Service),” [Online]. Available: https://en.wikipedia.org/wiki/Steam_(service)
2. Apple, “Turicreate,” [Online]. Available: https://apple.github.io/turicreate/docs/userguide/recommender/choosing-a-model.html