

DSCI 351: Recommender System

Project Proposal

Author(s): Andrew To

Project Title: Steam Video Game Recommender System

Date: July 27, 2025

1. Project Description

Steam is a digital online game platform that offers nearly 30,000 games from AAA to indie and everything in-between. It is one of the best platforms for players and game creators/providers to find their best match games/customers.

To fulfill this mission, Steam needs an effective recommender system to suggest most suitable games for players, which improves the experience of both players and game providers/creators.

This project aims to build a recommender system for Steam using combination of user review, user & item data, item metadata and bundle data that can be successfully and effectively implemented for Steam's recommender systems.

2. Dataset Source

https://cseweb.ucsd.edu/~jmcauley/datasets.html#steam_data

3. Dataset Description

This data source has 5 datasets:

- Dataset 1: Review Data: Data of text review from users
 - o 25799 samples
 - o Features:
 - User_id
 - User_url
 - Reviews
 - Item_id
 - Funny
 - Posted

- Dataset 2: User and Item Data: Data of user-item interaction
 - 88310 samples
 - Features:
 - User_id
 - Items_count
 - Steam_id
 - User_url
 - Items
 - Item_id
 - Item_name
 - Playtime_forever
 - Playtime_2weeks
- Dataset 3: Review Data: More data of text reviews
 - 7793069 samples
 - Features:
 - Username
 - Hours
 - Products
 - Product_id
 - Page_order
 - Date
 - Text
 - Early_access
 - Page
- Dataset 4: Item metadata: Product features
 - 32135 samples
 - Features:
 - Publisher
 - Genres
 - App_name
 - Title
 - url
 - release_date
 - tags
 - discount_price
 - reviews_url
 - specs
 - ...

- Dataset 5: Bundle (Collection of games at cheaper price)
 - 615 samples
 - Features:
 - bundle_final_price
 - bundle_url
 - bundle_price
 - bundle_name
 - bundle_id
 - items
 - genre
 - item_id
 - discounted_price
 - item_url
 - bundle_discount

4. Recommender systems

I plan to build and test with different approaches using various types of RS and data mining algorithms to find the best recommender system. In general, it will be a hybrid recommender system that combines user profile (information extracted from the first 3 datasets using TF-IDF for text reviews), product features (using the item metadata dataset), and these inputs will be tested with different machine learning and deep learning models.

Also, I consider applying knowledge-based models using information extracted from text reviews to identify user needs, or analyzing their history of discounted games purchased to see whether they prefer cheap games