Game recommendation system
Project Proposal
Mihail Kenarov

1. Introduction

What is the goal of our recommendation system?

The goal of this project is to create a recommendation system that allows a user to find similar video games to the one they have chosen. This will be determined by some of the features of the choice they have made and show other games that they may find interesting to get as well.

Why would we create such a system?

The reason for creating this project is to make it easier for the “consumer” to find more suitable games which match their taste.

Who would be interested in this project?

People who enjoy playing different kinds of video games may find this project useful, for future recommendations based on their interests.

When will the project be constructed?

- Phase 1 – Research and planning
  - Research about the project will be carried out in the first couple of weeks of the semester, to get a better understanding of recommendation systems and what would be the most suitable approach in our case.
- Phase 2 – Development
  - During that time the data will be edited and structured properly for the selected algorithm, which will also be chosen during this time, based on what currently matches our expectations.
- Phase 3 – Delivery and evaluation
  - After the creation and testing of the project it will be evaluated to see if it matches up to our expectations.

How will the project be created?

We will first get a suitable dataset that would give us the necessary information to help us create such a system. Afterwords we will see which recommendation algorithm works best in our case and find a suitable model to train. Afterwards inferencing will be done as a way of testing out the project and seeing if the recommendations work as expected.

2.Domain understanding

We all know what video games are, but not many of us know what exactly a recommendation system is and what it consists of, which brought me to create research of my own.

Research questions:

What is a recommendation system?

A recommendation system is a sophisticated tool designed to analyze user behavior, preferences, and interactions with items to predict what the user might like or need. These systems are integral to many online platforms, including e-commerce websites, streaming services, and social media, where they help users discover new content, products, or services that align with their interests. The goal of a recommendation system is to enhance user experience by providing personalized suggestions, thereby increasing engagement, satisfaction, and potentially sales or conversions.

How does a recommendation system work?

Recommendation systems operate on the principle of predicting user preferences based on historical data. This data can include user ratings, browsing history, purchase history, and interactions with other users. The system analyzes this data to identify patterns and correlations that can be used to predict future preferences. Not only that but it also tracks the similarities between products that a user has been interested in.

What types of recommendation systems are there?

1. ***Content-Based Filtering*** - This type of recommendation system suggests items to users based on the characteristics or attributes of the items themselves. It analyzes the features of items that a user has interacted with in the past and recommends similar items. For example, if a user has watched action movies in the past, a content-based filtering system might recommend other action movies with similar themes or actors.

2. ***Collaborative Filtering*** - Collaborative filtering recommends items to users based on the preferences or behavior of similar users. It identifies users who have similar tastes or interests and suggests items that those users have liked or interacted with but the current user hasn't. Collaborative filtering can be further divided into two subtypes:

   2.1. ***User-based collaborative filtering:*** It finds users who are similar to the target user based on their interactions with items and recommends items that those similar users have liked.
   2.2. ***Item-based collaborative filtering:*** It identifies items that are similar to the ones the user has interacted with and recommends those similar items.

3. ***Hybrid Recommendation Systems*** - Hybrid recommendation systems combine multiple recommendation techniques to provide more accurate and diverse recommendations. By leveraging both content-based and collaborative filtering methods, hybrid systems can overcome the limitations of individual approaches and offer improved recommendation quality.

4. ***Matrix Factorization*** - Matrix factorization techniques model the relationship between users and items by decomposing the user-item interaction matrix into lower-dimensional matrices. These methods are particularly useful when dealing with sparse data, such as user-item interactions in large-scale recommendation systems.

5. ***Context-Aware Recommendation*** - Context-aware recommendation systems take into account additional contextual information, such as time, location, device, and user activity, to provide more personalized recommendations. By considering the context in which recommendations are made, these systems can offer more relevant and timely suggestions to users.

6. ***Demographic-Based Recommendation*** - Demographic-based recommendation systems recommend items to users based on demographic information such as age, gender, location, and occupation. These systems can tailor recommendations to specific user segments or demographics, enhancing the personalization of the recommendation process.

What are some machine learning algorithms used in recommendation systems?

1. **Collaborative Filtering Algorithms:**

    1.1. K-Nearest Neighbors (KNN): This algorithm finds users or items that are most similar to the target user or item and recommends items or users that those similar entities have liked or interacted with.

    1.2. Matrix Factorization: Techniques like Singular Value Decomposition (SVD) and Alternating Least Squares (ALS) are used to decompose the user-item interaction matrix into lower-dimensional matrices. These decompositions can then be used to predict missing ratings or to recommend items.

2. **Content-Based Filtering Algorithms**

    2.1. Cosine Similarity: This algorithm measures the cosine of the angle between two vectors to determine how similar they are. In the context of recommendation systems, it can be used to compare the content of items to a user's profile to recommend similar items.

    2.2. TF-IDF (Term Frequency-Inverse Document Frequency): This technique is used to reflect how important a word is to a document in a collection or corpus. It can be used to represent items in a way that highlights their most important features, which can then be used to recommend similar items.

    2.3. K-Nearest Neighbors (KNN) for Content-Based Filtering: In addition to its role in collaborative filtering, kNN can also be effectively used in content-based filtering. By calculating the similarity between items based on their content features, kNN can recommend items that are most similar to the target item. This approach is particularly useful when the content features are numerical or can be easily compared using a distance metric.

3. **Hybrid Algorithms**

Hybrid Recommender Systems: These systems combine collaborative filtering and content-based filtering to leverage the strengths of both approaches. They can provide more accurate recommendations by considering both the similarity between users and the similarity between items.

4. **Deep Learning Algorithms**

    4.1. Neural Networks: Deep learning-based recommendation systems, such as those using Convolutional Neural Networks (CNNs) for image-based recommendations or Recurrent Neural Networks (RNNs) for sequence-based recommendations, have shown great promise in capturing complex patterns in data.
    4.2. Autoencoders: These are used for dimensionality reduction and can be particularly effective in recommendation systems where the data is high-dimensional. They can learn to represent items in a lower-dimensional space, making it easier to identify similarities.

5. **Ensemble Methods**

Ensemble Recommendation Systems: These systems combine predictions from multiple recommendation algorithms to improve the overall recommendation quality. Techniques like bagging and boosting can be used to combine the strengths of different algorithms.

6. **Graph-Based Algorithms**

Graph Neural Networks (GNNs): These are used in recommendation systems to model the relationships between users and items as a graph. GNNs can capture the complex relationships between entities, making them suitable for recommendation tasks.

Who is most affected by the usage and creation of recommendations systems?

Users

Users are the primary beneficiaries of recommendation systems. They receive personalized recommendations that help them discover new products, content, or services that match their preferences and interests. By receiving relevant suggestions, users can save time, make informed decisions, and enhance their overall experience on platforms such as e-commerce websites, streaming services, social media platforms, and more.

Businesses

Businesses that implement recommendation systems benefit from increased user engagement, retention, and conversion rates. By providing personalized recommendations, businesses can improve customer satisfaction, loyalty, and revenue generation. Recommendation systems also enable businesses to gather valuable insights into user behavior and preferences, which can inform marketing strategies, product development, and inventory management.

Advertisers

Advertisers benefit from recommendation systems through targeted advertising opportunities. By analyzing user data and preferences, recommendation systems can deliver personalized ads to users who are more likely to be interested in the advertised products or services. This targeted approach can improve ad relevance, click-through rates, and return on investment for advertisers.

Where are recommendation systems used?

There are many different platforms and places where the recommendation systems are used for a variety of purposes, having largely affect the market. Mainly they are used in the online sphere to help customers find things that are of similar interest to the things they like. Here are some examples of where the recommendations systems are used:

E-commerce and Retail

- Online Shopping Platforms: Amazon, eBay, and Alibaba use recommendation systems to suggest products based on a user's browsing history, past purchases, and the purchases of similar users.
- Clothing and Fashion Retailers: Stores like ASOS, Zara, and Nordstrom use recommendation systems to personalize clothing and accessory suggestions for their customers.
- Electronics Retailers: Companies like Best Buy and Newegg recommend products based on customer reviews, product specifications, and customer purchase history.

Streaming Services

- Music Streaming Platforms: Services like Spotify, Apple Music, and Pandora use recommendation systems to suggest songs, playlists, and artists based on a user's listening history and preferences.
- Video Streaming Platforms: Netflix, Hulu, and Disney+ recommend movies, TV shows, and series based on a user's viewing history, ratings, and the viewing habits of similar users.

Social Media

- Facebook and Instagram: These platforms use recommendation systems to suggest friends, groups, and content that a user might find interesting based on their interactions and the interactions of similar users.
- LinkedIn: Recommends job postings, companies, and networking opportunities based on a user's professional background, skills, and interests.

Online Advertising

- Google Ads: Google uses recommendation systems to suggest ads to users based on their search history, website visits, and the content of the websites they visit.
- Amazon Advertising: Amazon uses recommendation systems to suggest products to advertisers based on the products that similar users have viewed or purchased.

Travel and Hospitality

- Airbnb: Recommends accommodations based on a user's search history, past bookings, and the preferences of similar users.
- Booking.com: Suggests hotels, flights, and car rentals based on a user's search history, past bookings, and the preferences of similar users.

News Aggregators

- Google News: Recommends news articles based on a user's reading history and the reading habits of similar users.
- Reddit: Recommends subreddits and content based on a user's browsing history and the browsing habits of similar users.

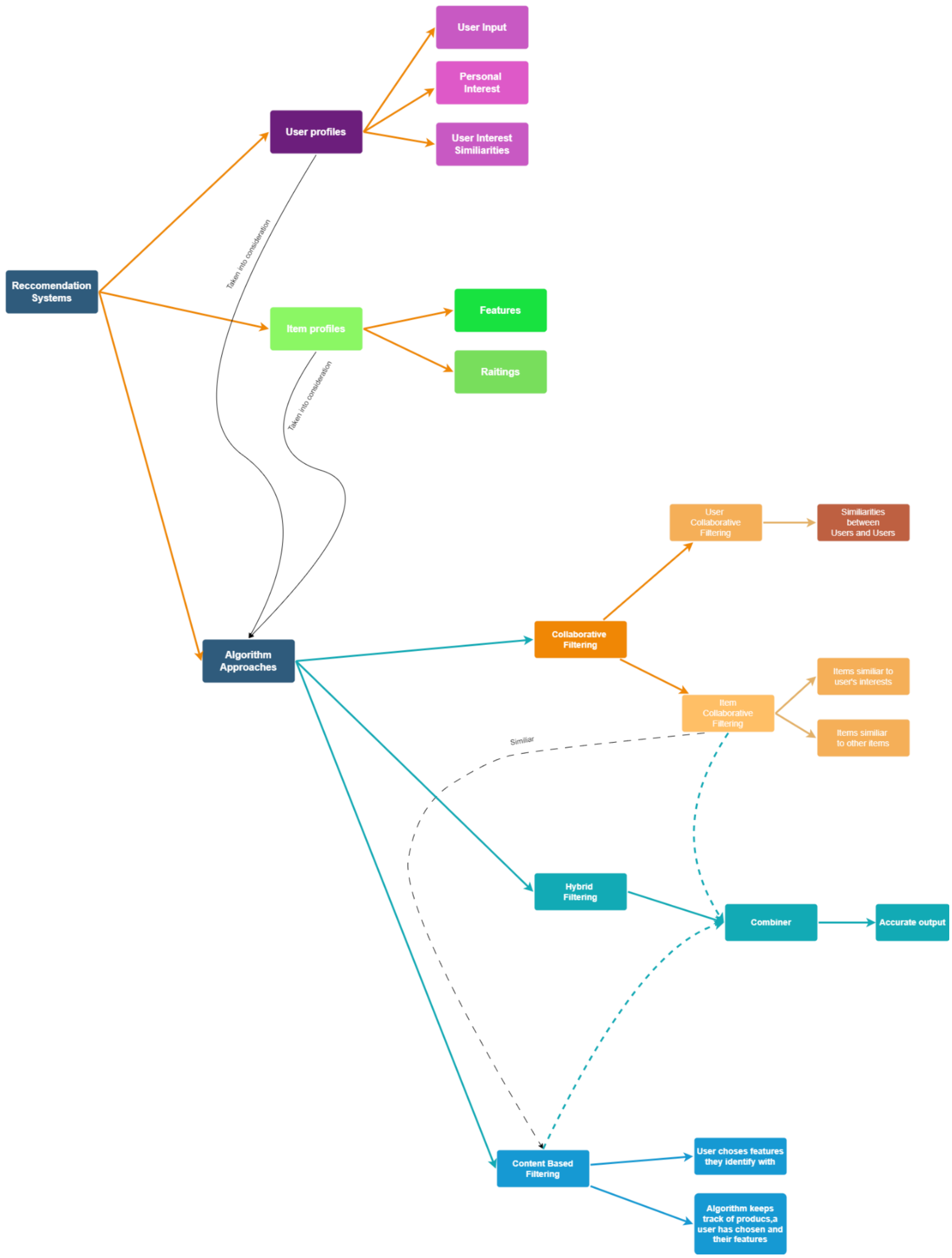
Education and Learning Platforms

- Coursera, Udemy, and Khan Academy: Recommend courses, tutorials, and learning paths based on a user’s past enrollments, ratings, and the enrollments of similar users.

Healthcare and Wellness

- Fitbit and MyFitnessPal: Recommend workout plans, diet plans, and health tips based on a user's activity data, dietary habits, and the habits of similar users.

As we can see, from the small examination of the given spheres of interests in the world, we can comfortably say that recommendation systems are big factor in the way that businesses expand and also the users are experiencing used to seeing them everywhere – no matter is if it is a digital or an actual holdable product.



Reccomendation Systems
User profiles
User Input
Personal Interest
User Interest Similiarities
Item profiles
Features
Raitings
Taken into consideration
Taken into consideration
Algorithm Approaches
Collaborative Filtering
User Collaborative Filtering
Similiarities between Users and Users
Item Collaborative Filtering
Items similar to user's interests
Items similiar to other items
Similiar
Hybrid Filtering
Combiner
Accurate output
Content Based Filtering
User choses features they identify with
Algorithm keeps track of producs,a user has chosen and their features

3. Data Sourcing

Objective

Considering that the mission of the project is to create a recommendation system for video games, which will help gamers find more alternatives that they find interesting.

Data Requirements

The dataset that has been selected is full of text data as well as some numerical data that will be useful to our project. Also, all of the columns have been labeled which will make it easier for us to understand it.

Data Sources

Publicly available dataset - https://www.kaggle.com/datasets/asaniczka/video-game-sales-2024

Actual source of the dataset - https://www.vgchartz.com/

The site seems to specialize in showing different kinds of data that would be interesting for gamers. We are shown variety of charts and articles about different topics. It varies from hardware to software as well. It does seem quite trustworthy and shows a lot of information that can be useful to the users. When it comes to the data – it has been collected in 2024, so we can comfortably say that it is recent and quite relevant.

Data Legality and Ethics

It is publicly available, and we have also been given a source from where it has been scraped from.

Data Diversity

(0) Title: The name of the game - Text

(1) Console: The console on which the game is played on- Text

(2) Genre: The genre of a game - Text

(3) Publisher: The game publishers which could be considered 'The big names' of the industry - Text

(4) Developer: The studio that worked on the creation of the game - Text

(5) Critic score: The score that is given to a game from a certain agency (like IGN) - Number

(6) Total sales: The number of times the game has been sold worldwide - Number in millions

(7) NA sales: The number of times a game has been sold in North America - Number in millions

(8) Japan sales: The number of times a game has been sold in Japan - Number in millions

Version Control

There are some missing fields like the ratings of a game on a specific console, but it can be fixed with some after some data cleaning. This means that some changes do need to be made to the chosen dataset.

A history of the project and the processing of the data will be kept on a Git repository in case of an incident and as a way for a version control.

The link to the git repository:

- https://git.fhict.nl/I509460/video-game-reommendation.git

Iterative Process

The model will be check for its accuracy continuously and depending on the results, more processing of the data is going to be done.

4.Analytic Approach

The target is to give the user the titles of multiple games which can interest them for them to buy. This will be done by finding similar games to the one they have selected and based on the features of the game, find other ones that will be similar to it.

The problem is that the model is a bit of a hybrid between a classification issue (currently the genre) as well as some would say it is a regression issue (because of the rating).

The chosen model for the task currently is Nearest Neighbors (kNN) and we will try to determine the success of it based on the accuracy. We will be using kNN as a starting point, since it is a great algorithm for the creation of a system that is of the classification type, while also giving us the desired effect. This will be a good choice since we are currently going for a content-based filtering system.