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
    - o 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
    - o 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
    - o After the creation and testing of the project it will be evaluated to see if it matches up to our expectations.

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 the first question:

What is a recommendation system?

A recommendation system is a technology-driven tool that suggests products, services, or content to users based on their preferences, behavior, and interactions, or similar to a product that they have chosen.

Taking this into account, it becomes clear that we also need a user and some of their qualities as well as the features of the items they interact with.

What types of recommendation systems are there?

There 3 main types of recommendation systems:

- Content-based filtering
- Collaborative Filtering
   o   User-User collaborative filtering
   o   Item-Item collaborative filtering
- Hybrid system

What are some are some approaches and machine learning techniques that are used when creating a recommendation system?

- K-nearest neighbors
- Cluster analysis
- Decision trees
- Bayesian Classifiers
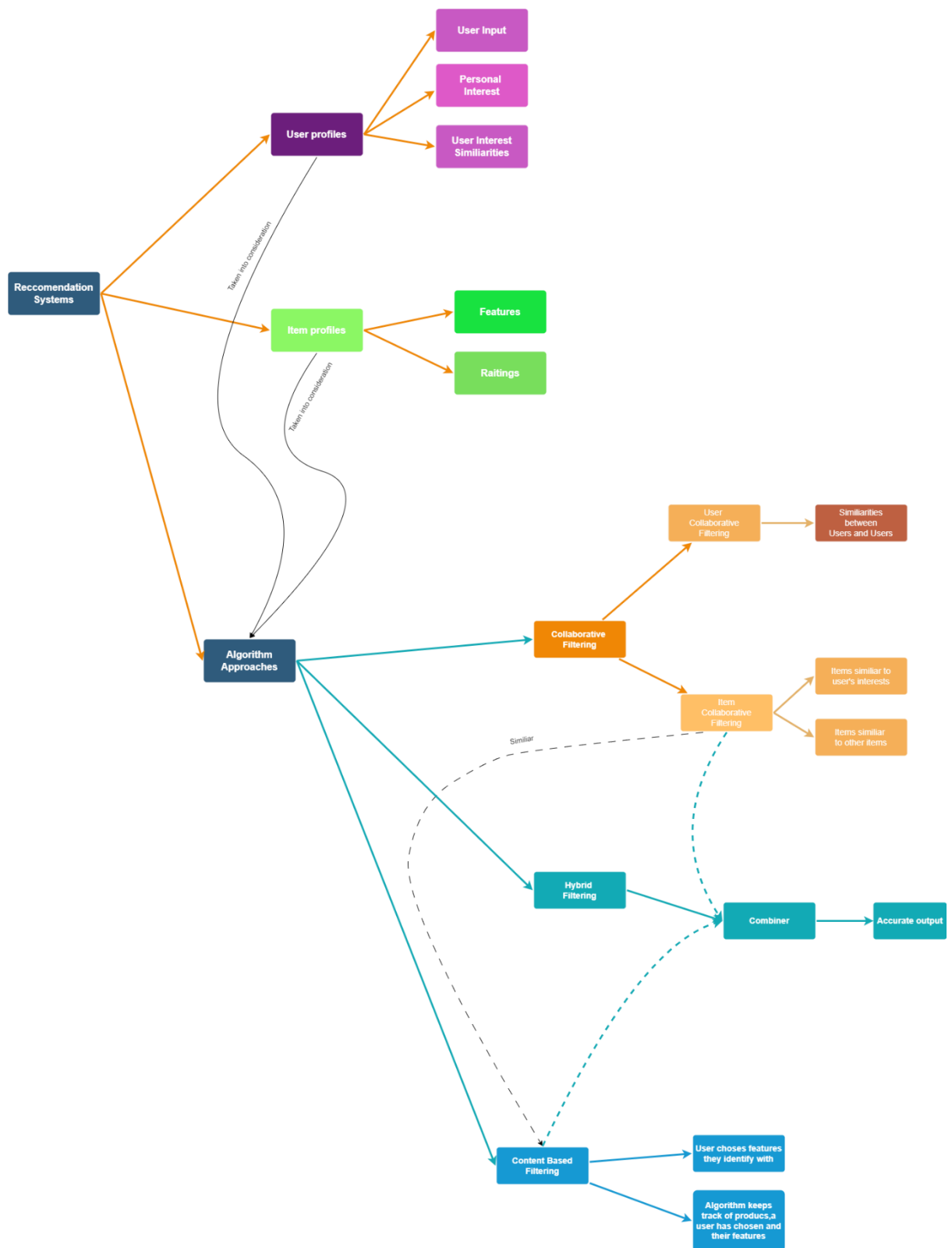
3. Data Sourcing

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

Data Legality and Ethics

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

Data Diversity

The data contains the different genres of the games, as well as publishers, ratings, sales and other features.

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.

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.

The problem is that the model is a bit of a hybrid between a classification issue (publisher and 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.