Videogame Recommender System

Big Data Computing Project
A.Y. 2022-2023

Lucia Fores 1836451

Roadmap

I - Background and Goals

II - Data Collection

III - Data Analysis

IV - ALS Training

V - ALS Analysis

VI - Clustering

VII -Clustering Analysis

VIII - Demo

I - Background

The project goal is to create a Videogame Recommender System based on the Steam platform.

Steam is, at the moment, the most comprehensive and widely used videogame store on PCs that can count on a huge and active community

I - Goals

1. Create a Collaborative Filtering Recommender System based on the reviews left by the Steam user on the platform

2. Inspect a new Recommender System approach based only on the plot similarities of the games

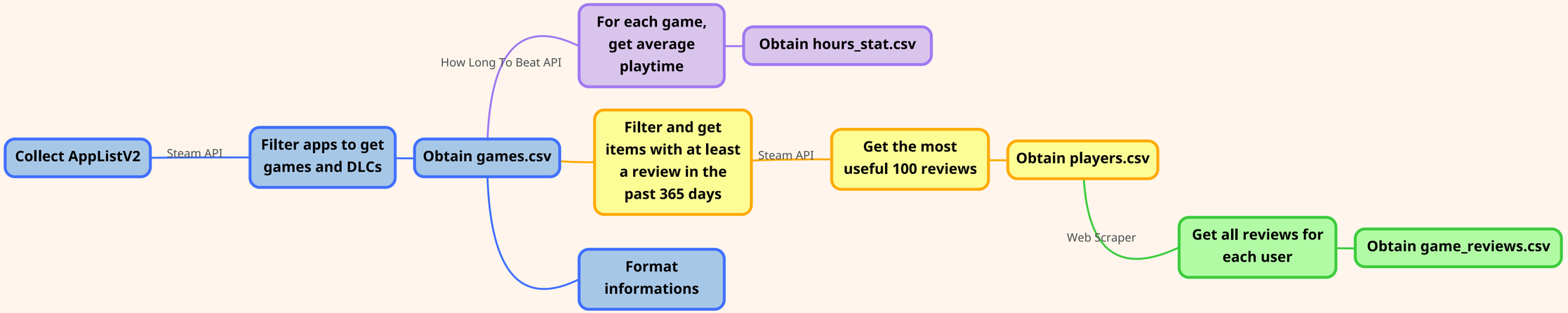
II - Data Collection

The data collection was divided in:

- Game data collection
- User IDs collection
- Reviews collection
- Hours statistic collection

Different technologies has been used during the collection like APIs and Web Scrapers

II - Data Collection



III - Data Analysis

Tables Obtained

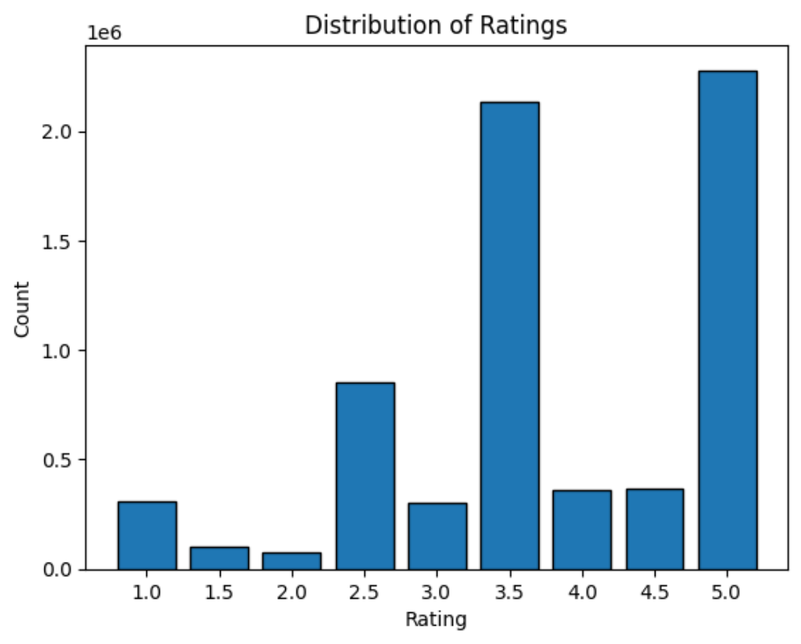
Data arranged in two tables:

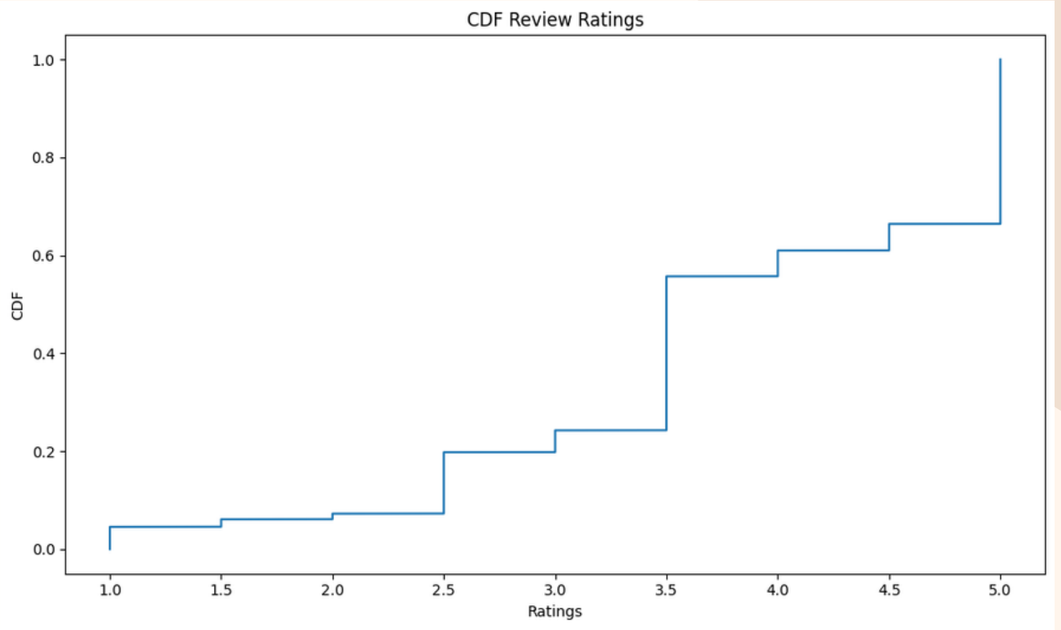
- games.csv, containing all the information about the videogames in the store
- game_reviews.csv, containing a sample of the reviews (about 7 million) that the users left on the collected games

Translation

Translated the positive/negative reviews into a larger system of ratings (ranging from 1 to 5) considering the time that the user played the game and the average time needed to complete the game

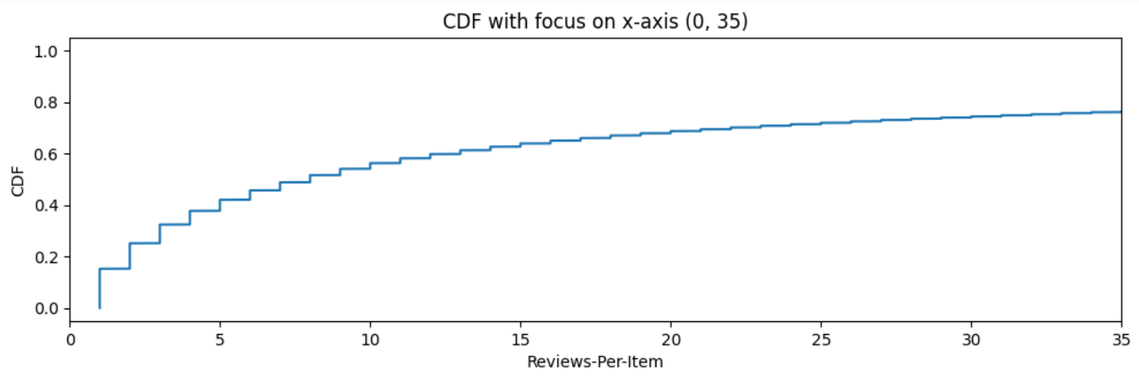
III - Data Analysis

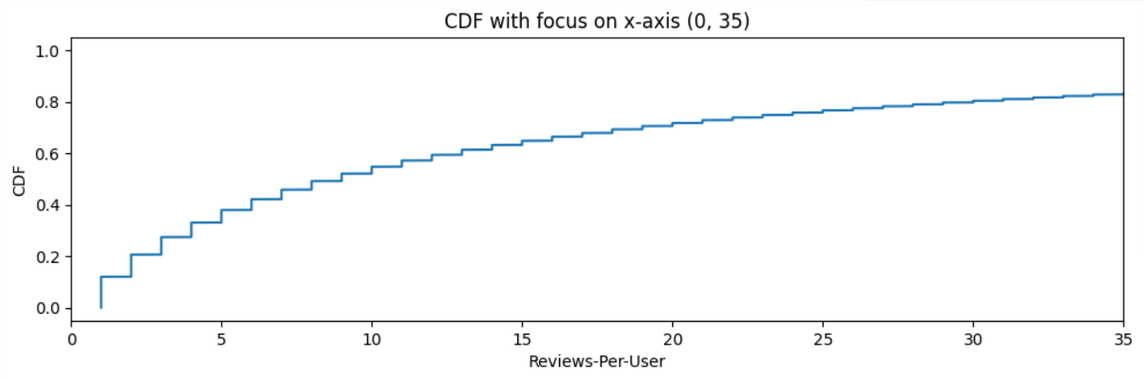




Got a sample of reviews that, after the translating process to ratings, showed a majority of positive ratings (ranging from 3.5 to 5), this type of ratings in fact represented about the 75% of the total sample

III - Data Analysis





Most of the videogames in the sample have 35 or less reviews and that most of the users left 35 or less reviews each meaning that the matrix representing the interaction between users and items in the system will be very sparse

IV - ALS Training

1. Created a Collaborative Filtering Recommender System Model Based specifically with an Alternate Least Square (ALS) model

2. Evaluated the model with the Root Mean Square Error (RMSE) error measure and trained the model through a cross validation

3. Used a grid search in order to find the best values for the hyper-parameters (rank, number of iterations and regularization parameter)

4. The best model obtained, with RMSE equal to 0.944, had:
   - Rank: 15
   - regParam: 0.1
   - maxIter: 20

V - ALS Analysis

Analysis Performed

In the analysis performed a function to compute score has been created. With the function it is possible to compute Precision@K, Mean Average Precision (MAP), Recall@K and Average Normalize Discounted Cumulative Gain (avg_NDCG)

Results

The trained model resulted to have:

- RMSE: 0.944
- Precision@K: 0.000033
- MAP: 0.000029
- Recall@K: 0.000023
- avg_NDCG: 0.000033

VI - Clustering

1. With the goal to create clusters in which the items inside are videogames with similar plots, created a K-Means model

2. Carried out the clustering over two different type of plots present in games.csv, called 'about_the_game' (longer) and 'short_description' (shorter)

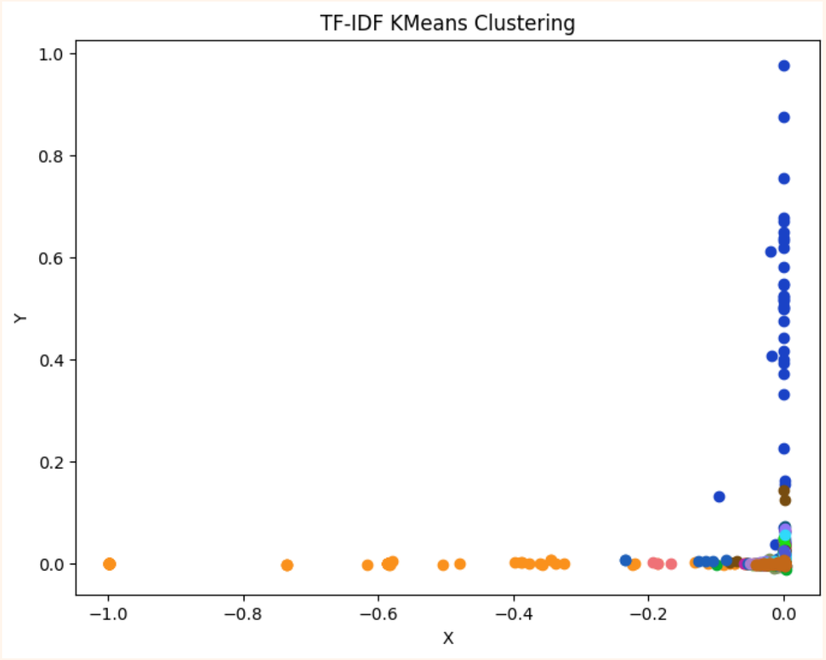
3. Performed a typical NLP pipeline in order to pre-process the texts to be processed by the cluster algorithm. The PCA has been then carried out in order to visualize the clusters

4. Tried different hyper-parameters values. The best configuration were:

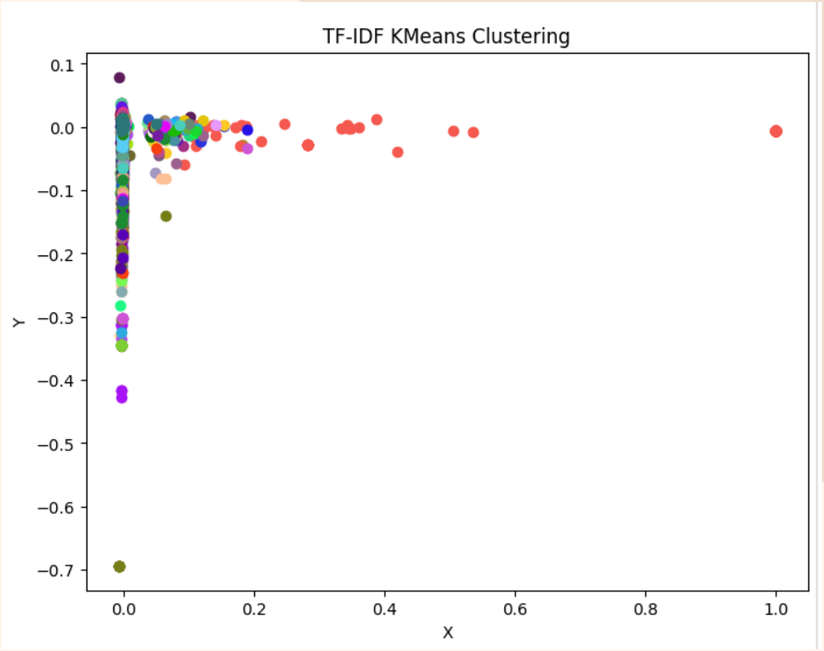
'about_the_game'
- K: 50

'short_description'
- K: 300

VII - Clustering Analysis



‘about_the_game’



’short_description’

The performances of the clustering has been evaluated thanks to the Silhouette metric and the Within Cluster Sum of Squared Distances.

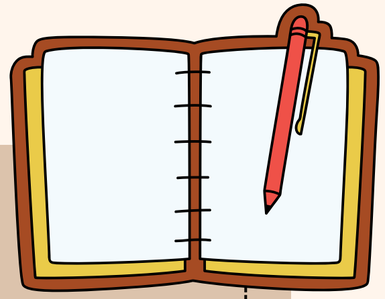
The algorithm could not find a distinct separation among the items.

VII - Demo

Web application in which it is possible to require either recommendation for a specific user or recommendation over a specific videogame.

The demo was created with Streamlit

Future Developments



- Increase the dataset
- Add the textual feedback of the reviews
- Use the information about the achievements collected by a user in a game
- Increase the performance of the models

Thanks for the attention!

Lucia Fores 1836451