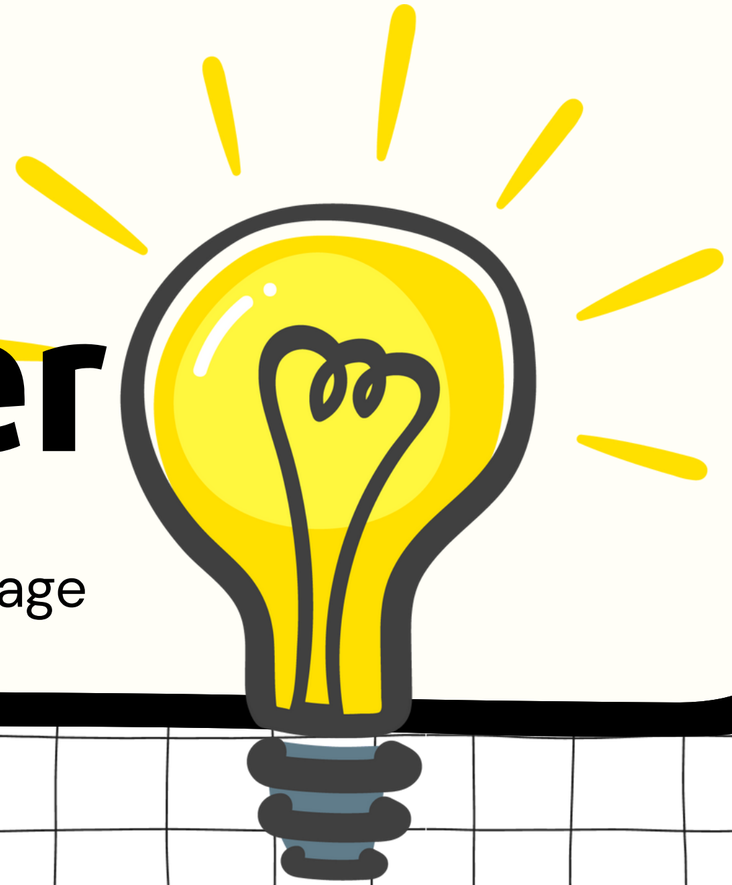
Game Recommender



Machine Learning Fortnight 2023 - Team Package

Contents

Who are we?

Cezar Bulancea

Creating as much leakage as possible

Alexander Müller

Hyperparameter tuning on test data

Problem Statement

- This year's task was to build the best algorithm to **predict if a user will like a certain game or not**
- Multiple datasets were provided:
  - training data which contained information about the reviews left for certain games by different users
  - game metadata which consisted of information about several games (around 1/3 of the games also present in the training data)
  - testing data was used for computing the final predictions

Attempted methods

01 Collaborative filtering

02 Ensemble learning

03 Hybrid model

Base model

Simple but ineffective

Better, but harder...

Data Analysis

- Training dataset features used:
  - steamid, appid, voted_up
- Meta dataset features used:
  - the total number of recommendations
  - publishers + categories + genres
- Feature engineering
  - we extracted the year of publishing of each game from the release date
  - we created a ‘soup’ feature = publishers + categories + genres
  - we also created a weighted vote score using a formula found on the internet

Results

- **Collaborative filtering**
  - F-score of approximately **0.89**
  - We used the **FastAI** library to implement our model
  - Used **weight decay** as a regularization technique
- **Hybrid model**
  - We combined the collaborative filtering approach with a simple recommender, but the approach was ineffective
  - Using a more complex content-based model combined with collaborative filtering might lead to significant results

What we learned?

- Designing recommendation systems is no easy task :)
- The quality of the data is one of the most important factors when it comes to building useful machine-learning models
- It can get frustrating when trying a lot of ideas that end up being trash ⟶ a strong mindset can be useful
- We are certain of one thing though: nothing was in vain, we just discovered almost 100 hundred ways that don’t work :)

Conclusions

Difficult process

More challenging

A lot of fun!

Thank you!