Recommender Systems Project

Masina A & Moyeni N

01

Background

Table of Contents

About the Project

In today's technology-driven world, recommender systems are socially and economically critical for ensuring that individuals can make appropriate choices surrounding the content they engage with on a daily basis. One application where this is especially true surrounds movie content recommendations; where intelligent algorithms can help viewers find great titles from tens of thousands of options. Recommender systems are used by many streaming companies to recommend content to customers based on their viewing preferences. This project aims to use a similar system to build a recommender model to acurately predict how a user will rate an anime title they have not yet viewed.

Dataset Features:

4 data sets

- Anime.csv
- Test.csv
- Train.csv
- Submission.csv

The Team

ALEX MASINA

Data Scientist

NTHABISENG MOYENI

Data Scientist.

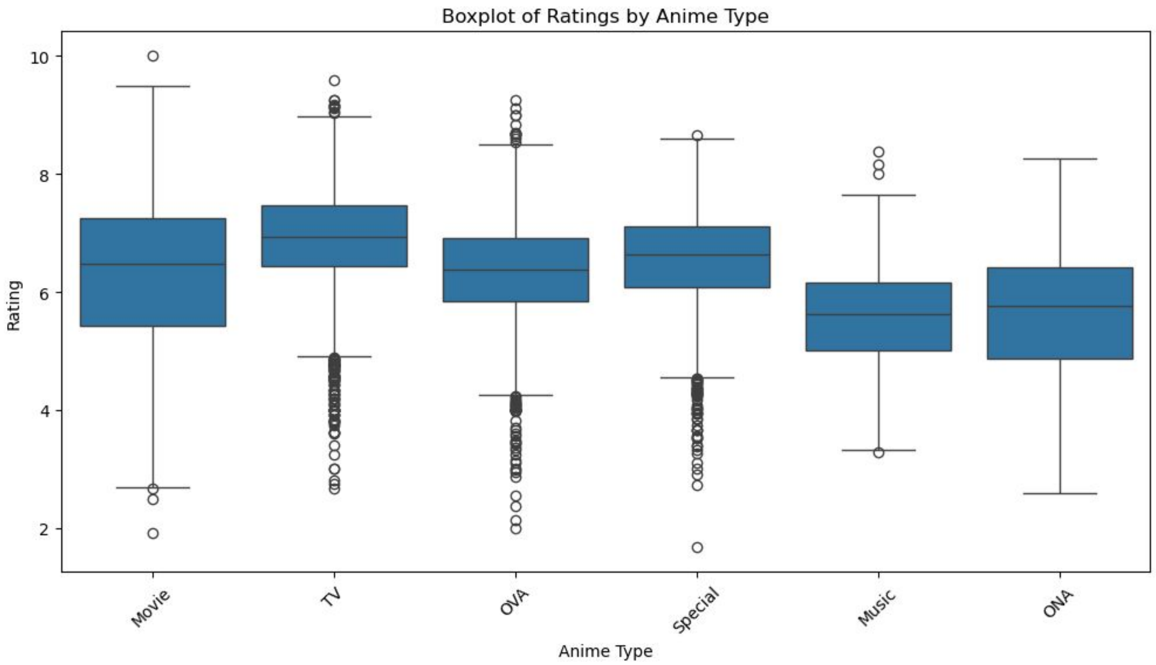
02

About the Data

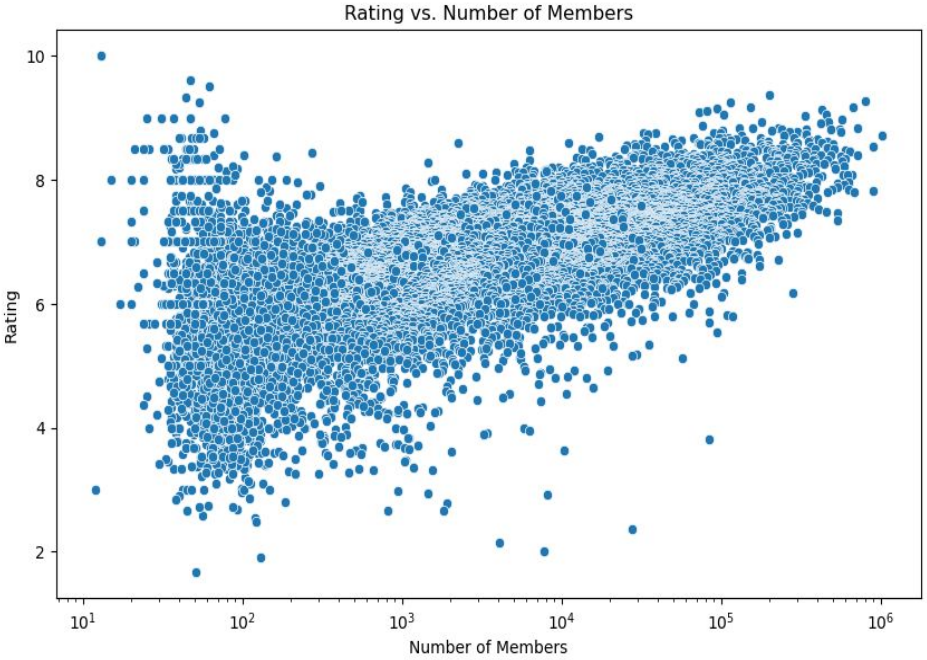
About the data

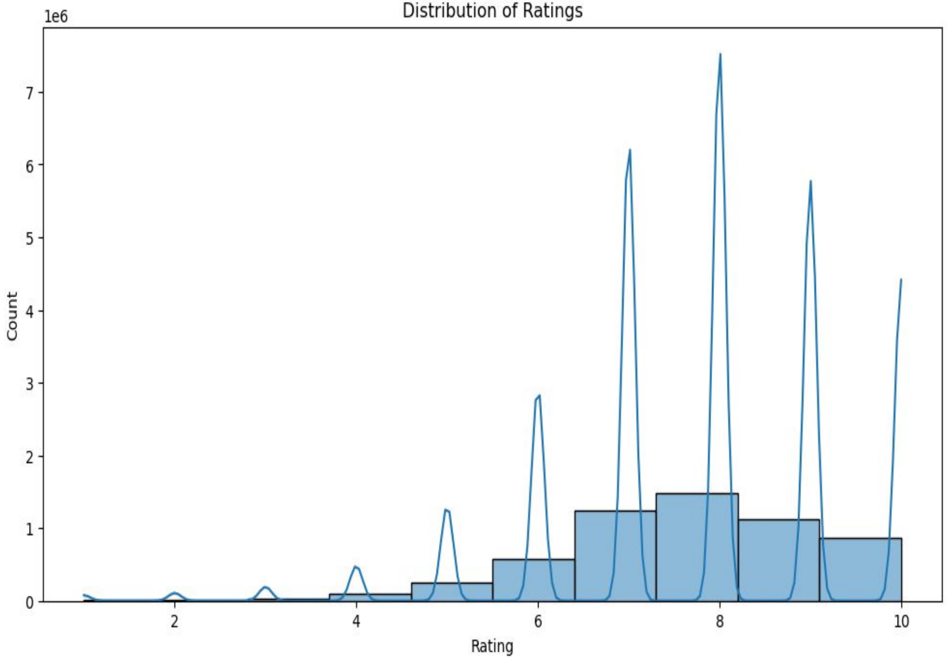
test_ =
https://github.com/AlexMasina/Unsupervised-Learning-Project/blob/main/test.csv
train_ =
https://github.com/AlexMasina/Unsupervised-Learning-Project/blob/main/train.csv

- The ratings on the training data do not seem to show any biaseness



About the data





03

Model Building & Validation

Model Build

We looked at four different builds

- 1. Linear Regression
- 2. Logistic Regression with combined TF-IDF and genre features
- 3. Logistic regression with combined all features
- 4. XGBoost with additional features

The model with additional relevant features seems to have a lower RMSE, we will use this model in our Streamlit app

Model Results

These RMSE (Root Mean Squared Error) results indicate the performance of different models and feature combinations in predicting the anime ratings. Here's a quick analysis of each result:

* Baseline RMSE: 0.20273655031875265

This is the RMSE of the initial model without any combined or additional features.

* RMSE with Combined Features: 0.1933,

By combining features such as genre, type, and normalized episodes, the model's performance improves, leading to a lower RMSE.

* RMSE with Additional Features: 0.1908 ,

Including more relevant features further enhances the model's accuracy, resulting in an even lower RMSE.

RMSE with XGBoost and Additional Features: 0.33167,

Surprisingly, the RMSE with XGBoost and additional features is higher compared to other models. This could be due to several reasons, such as overfitting or the model's sensitivity to hyperparameters.

04

Streamlit App Deployment

Streamlit App

Streamlit is an open-source framework designed to create interactive, web-based applications for data science and machine learning projects. It allows developers and data scientists to turn their Python scripts into shareable web apps with minimal effort.

- Script can be found on Reccomender_App.py

Streamlit App

Questions?

SIYABULELA

THANK YOU

RE A LEBOHA

BAIE DANKIE

NDO LIVHUWA