SeamCast Recommendation System

Background and breakdown

Ian Musau
DS-FT08

Overview

SeamCast is a planned movie streaming service that will also allow separate users to stream the same film in sync. They are in need of a robust movie recommendation system that users can enjoy together.

Business Problem

Given that the system is not only a streaming platform, but will also allow multiple users to stream the same film concurrently to watch together, SeamCast is in need of a system that will:

- Find the best suited movies for a given user
- Recommend movies that will be enjoyed by all parties in the group
- Identify different subsets of users and cater to their tastes

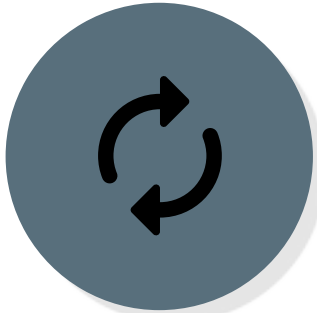
Limitations

- The system was built using less than 1% of the data available, so it is not as robust as one built on the entire dataset.
- Due to computation limits, some features of the movies such as cast and directors were not used. They would enhance the system’s prediction ability.

System Overview

Cluster

Users will be clustered into groups according to their tastes.



Recommend

The most popular movies from a given cluster are determined.

Predict

From recommended movies, models predict whether a user is likely to enjoy a movie or not.

Cluster model performance

Model type: KNN
Job: Predict which cluster a user would belong to

Accuracy: 99.55%

	precision	recall	f1-score	support
0	0.99	0.99	0.99	696
1	1.00	1.00	1.00	2163
2	1.00	1.00	1.00	4962
3	0.99	0.98	0.99	1522
4	0.98	0.99	0.99	486
accuracy			1.00	9829
macro avg	0.99	0.99	0.99	9829
weighted avg	1.00	1.00	1.00	9829

Observations

The model will group newer users into the appropriate clusters accurately

Prediction model performance

Model type: Decision tree forest
Job: Predict whether a specific user will like a movie

Training score:67.24%
Test score: 67.36%
Mean Cross Validation Score for Random Forest Classifier: 67.10%

Notes

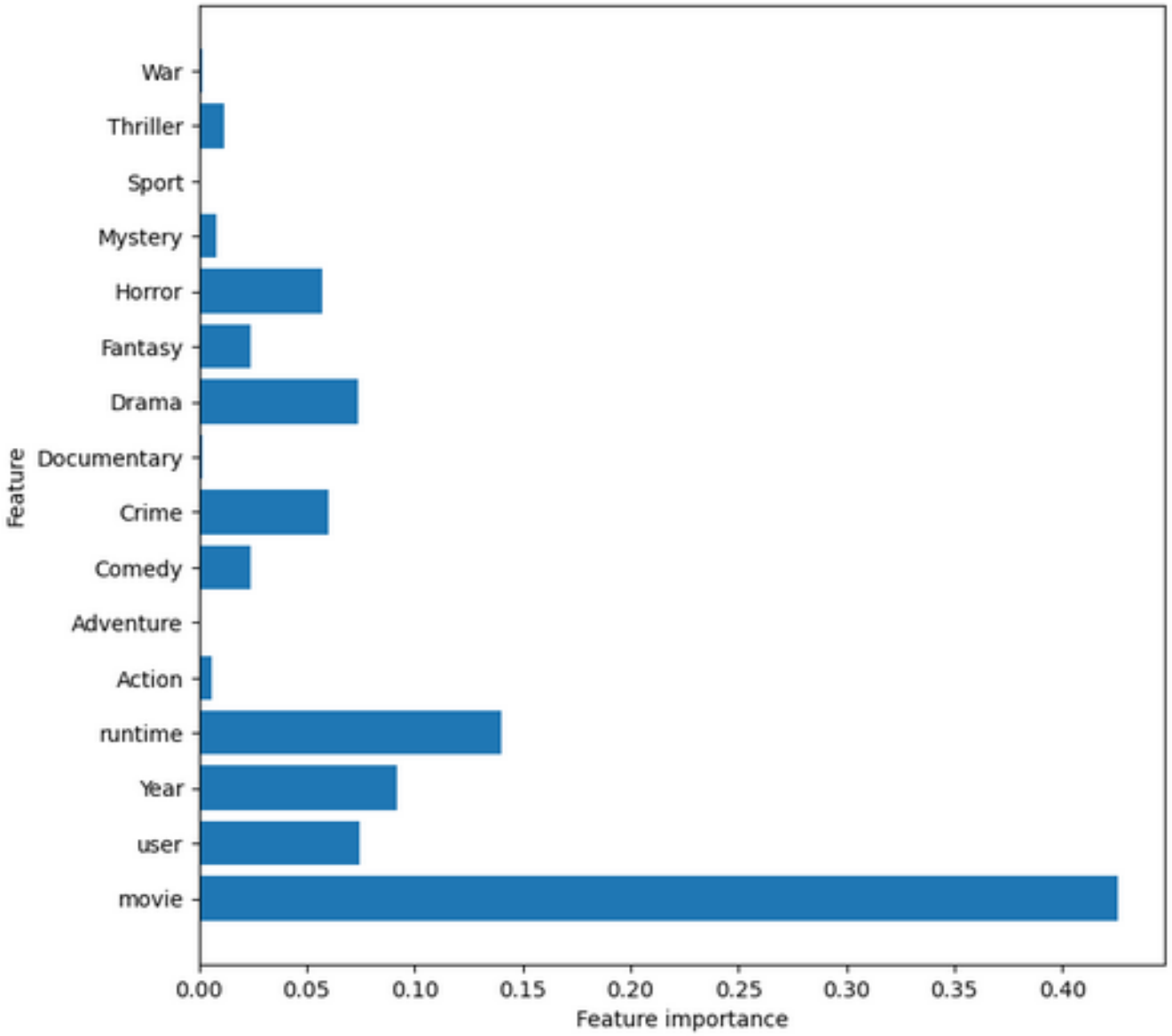
- With 67% accuracy the model falls just short of a serviceable score of 70%. This is due to the fact that only a small subset of the data (less than 1%) was used
- The proposed system would do well for watch parties by finding movies commonly popular in all users' clusters
- With added computing power, more features of the movies could be added to the model. These include directors, cast, language and key themes.

Prediction Model Feature Importance

Model type: Decision Tree Forest
Job: Predict whether a user will enjoy a movie

Observations

- Runtime of the movie had a high impact on the ratings of the viewers
- Drama, crime and horror genres seem to illicit the strongest reaction from users



Conclusion

- The use of clustering and prediction classification is a powerful solution to the recommendation problem
- With the aforementioned suggestions, the final version of the system would rival that of any giant in the streaming space
- SeamCast would do well to fund further development of the proposed solution to ensure domination of the streaming sector

Thank You

Any questions welcome