Movie Recommendation Systems

# **Conrad Ibanez Spring 2021 https://cvibanez.github.io/**

# Which Domain?

The domains for this project cover both e-commerce and entertainment as the topic involves movie recommendations for customers. The references are:

1. <https://www.kaggle.com/rounakbanik/the-movies-dataset> – Kaggle dataset which is the primary dataset for this project.
2. <https://www.kaggle.com/rounakbanik/movie-recommender-systems> - Recommender system project by the same user for the dataset
3. <https://www.kaggle.com/ibtesama/getting-started-with-a-movie-recommendation-system> - Another project for the dataset
4. <https://grouplens.org/datasets/movielens/latest/> - The full dataset on which the Kaggle dataset is based.
5. Bobadilla J., Ortega F., Hernando A., Gutierrez A., “Recommender systems survey,” Knowledge-Based Systems 46 109-132, 2013. – Discusses different research and approaches to recommendation systems.
6. Bennett J., Lanning S., "The Netflix Prize," Proceedings of KDD Cup and Workshop 2007, 2007.- Competition that sparked the interest for this project.
7. Covington P., Adams J., Sargin E., "Deep Neural Networks for YouTube Recommendations," Proceedings of the 10th ACM Conference on Recommender Systems, 2016. – Research paper for neural network
8. Harper F.M., Konstan J., "The MovieLens Datasets: History and Context," ACM Trans. Interact. Intell. Syst. V, N, Article XXXX, 2015. – More details on the MovieLens dataset
9. Koren Y., Bell R., "Advances in Collaborative Filtering," Recommender Systems Handbook, 2011. – Discusses collaborative filtering
10. Smith B., Linden G., “Two Decades of Recommender Systems at Amazon.com,” IEEE Internet Computing, 2017. – Discusses Amazon deployment of their recommendation systems.

# Which Data?

The dataset for this project is The Movies Dataset found on Kaggle- <https://www.kaggle.com/rounakbanik/the-movies-dataset>. This dataset contains metadata for the Full MovieLens Dataset (<https://grouplens.org/datasets/movielens/latest/>), which contains over 45,000 movies and over 26 million ratings from more than 270,000 users. The Movies Dataset contains seven csv files of data out of which two files are subset data files.

# Research Questions? Benefits? Why analyze these data?

As more and more businesses leverage data to improve their bottom line, one area that stands to benefit is e-commerce. People are now purchasing more items online as opposed to going into physical stores, and we have seen the impact of digital growth in the entertainment industry. The growth of online streaming companies led by Netflix and the downfall of brick-and-mortar stores, such as Blockbuster, are evidence of the shift in consumer behavior.

This project will analyze the data and discuss recommendation systems. It will investigate the different approaches that are used by recommendation systems and evaluate their performance. This project will address various questions. How do the different approaches work? What are the benefits versus the disadvantages? Does one approach perform better than the others? Is there a better method that can be created?

The benefits of deploying the best recommendation systems include increasing sales and improving the satisfaction of customers. If a company can sell more of its products that customers want, it is a win-win scenario for both parties. Hence, companies are now spending more to improve their marketing efforts by utilizing vast amounts of data.

# What Method?

There are many approaches in the design of recommendation systems. One common approach is collaborative filtering. Collaborative filtering makes recommendations based on patterns of ratings or usage. Another approach is content-based filtering where a system determines what a user would like by analyzing information describing the item while factoring in the user’s preferences. Other approaches for recommendations systems combine different aspects to form a hybrid approach. The different techniques used for recommendation systems include k- nearest neighbors, Bayesian networks, neural networks, and genetic algorithms.

# Potential Issues?

There could be some challenges with this project around the data, the different approaches, and different techniques. The data may not have enough information to implement different approaches for the recommendation system. Additionally, there is the question of how many of the different techniques can be used for the data.

# Concluding Remarks

We live in a world full of choices when it comes to entertainment and the programs we watch. With the growth online streaming, movies and shows are available at our fingertips. Recommendation systems have become a common feature in many e-commerce sites including streaming services. Users interact with these systems to have more personalized experiences and to explore new offerings. These systems are based on different algorithms and filtering methods, such as collaborative filtering and content-based filtering. They factor in different sources of information to provide recommendations that users will want. This project will analyze the different approaches and techniques for movie recommendation systems and evaluate their performance.