MOVIE LENS RECOMMENDATION SYSTEM

Carlos Govea

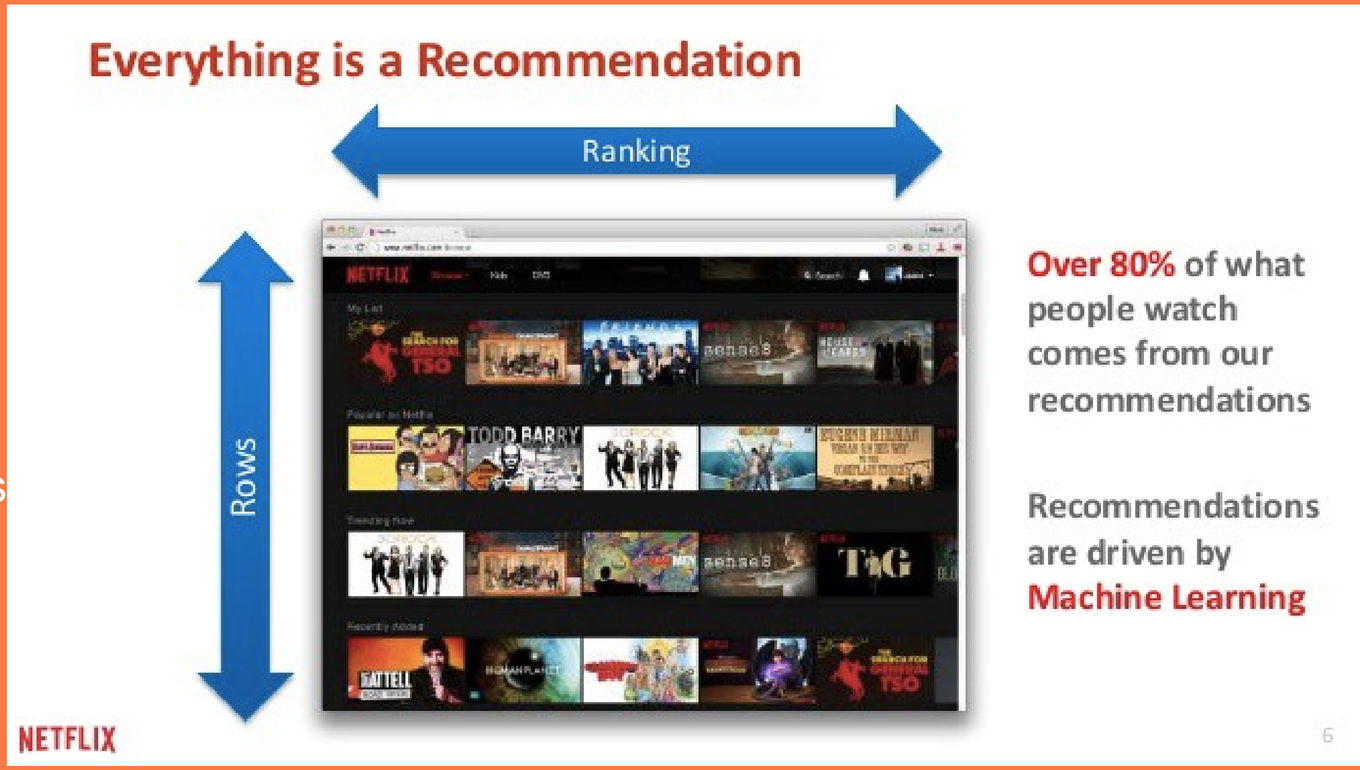
linkedin.com/carlos-govea-GG

github.com/CarlosGG18/Reccomendation-System

Business Understanding

Massive options pool make making decisions hard. Machine learning recommendation systems make it easy to:

- Increase retention rate
- help businesses to understand the viewing habits and preferences of their customers
- tailor marketing and content acquisition strategies



Recommendation Method

Recommender systems are electronic applications, the aim of which is to support humans in this decision making process

Based on the past behavior and preferences of users. It looks at how similar people have rated items in the past, and using that information to make recommendations to a new user.

	 Harry Potter	 The Triplets of Belleville	 Shrek	 The Dark Knight Rises	 Memento
	✔		✔	✔	
		✔			✔
	✔	✔	✔		
				✔	✔

Dataset

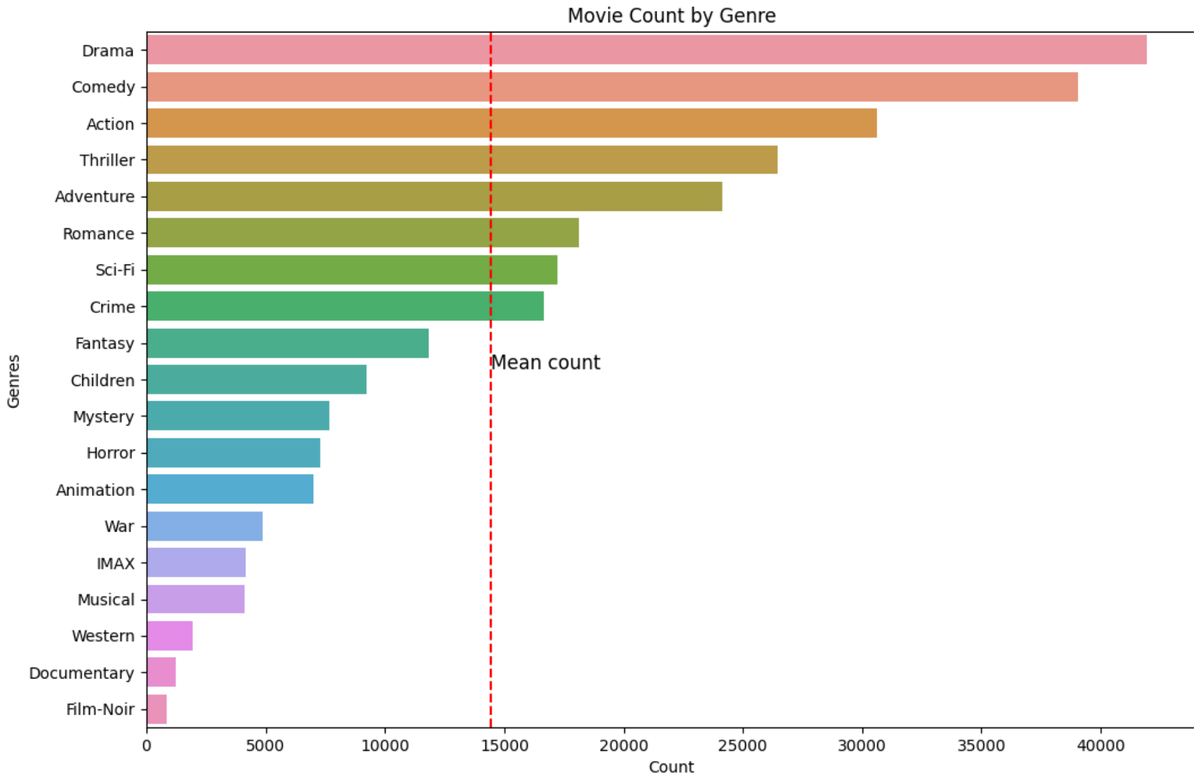
MovieLens dataset collected by University of Minnesota that contains User-ids, Movie-ids, Title, Genre, and ratings ranging from 1-5

Collaborative filtered recommendation systems only need UserId, MovieId, and Rating

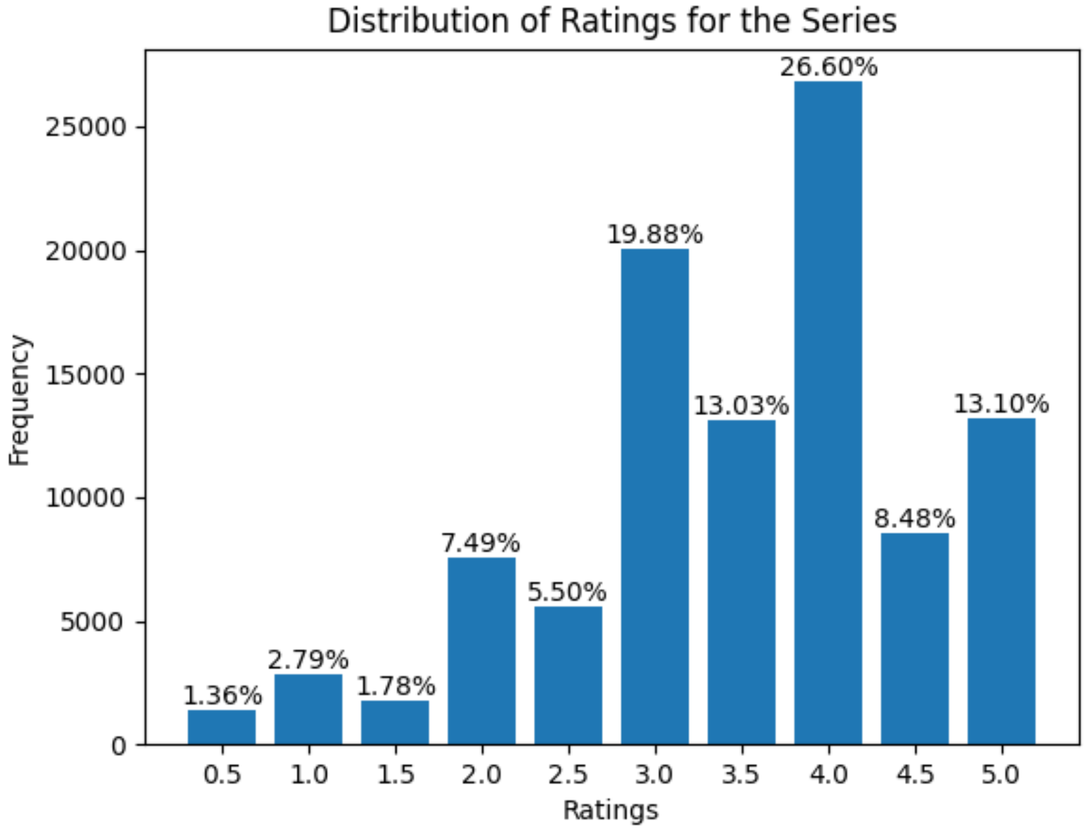
	userId	movieId	rating	title	genres	year
0	1	1	4.0	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	1995
1	1	3	4.0	Grumpier Old Men (1995)	Comedy|Romance	1995
2	1	6	4.0	Heat (1995)	Action|Crime|Thriller	1995
3	1	47	5.0	Seven (a.k.a. Se7en) (1995)	Mystery|Thriller	1995
4	1	50	5.0	Usual Suspects, The (1995)	Crime|Mystery|Thriller	1995

Exploratory Data Analysis

Top 4 genres dominate rated movies, may cause popularity bias



Users tend to be more discerning in their ratings, rather than consistently giving high or low ratings.



Model Results

My best performing model was the tuned KnnBaseline having the lowest RMSE and MSE overall

Other models like CoClustering and SVD may perform better with larger datasets

Model	RMSE	MSE
SlopeOne	0.90	.80
KnnBasic	.97	.75
KnnBaseline	**.68**	**.46**
SVD	.70	.66
CoClustering	.80	64

Example Recommendation

Select random user, in this case userId :[10]

Watched Movies

	userId	movieId	rating	title	genres	year
1119	10	296	1.0	Pulp Fiction (1994)	Comedy Crime Drama Thriller	1994
1120	10	356	3.5	Forrest Gump (1994)	Comedy Drama Romance War	1994
1121	10	588	4.0	Aladdin (1992)	Adventure Animation Children Comedy Musical	1992
1122	10	597	3.5	Pretty Woman (1990)	Comedy Romance	1990
1123	10	912	4.0	Casablanca (1942)	Drama Romance	1942

Recommended Movies

	movieId	estimated_rating	title	genres
0	3567	5	Bossa Nova (2000)	Comedy|Drama|Romance
1	187717	5	Won't You Be My Neighbor? (2018)	Documentary
2	100556	5	Act of Killing, The (2012)	Documentary
3	6818	5	Come and See (Idi i smotri) (1985)	Drama|War
4	3379	5	On the Beach (1959)	Drama

User 10 is a fan of Comedy, War, and Drama so the suggestions are inline with the users' preference

Problems Faced & Future Work

Cold start problem ~ Difficult to make recommendations for new users with no history; Incorporate better tuned models/ Neural networks

---

Try using different metrics like precision/recall to measure the proportion of relevant items among the total number of recommended

---

Additional models helps with limited diversity issues with only using collaborative methods.