Movie Recommendation System

Presented by: Jiarong Chen

Background

Business problem

Using the movie lens 100k data to analyze the movies' rating and make a movie recommendation system.

Outline

1. **Best Rating Movie**
   - Count the total movie rating and rating times
   - Sort the movies by mean rating
   - Determine the best movie base on mean rating and rating times
2. **Best Movie Genre**
   - Build a genre data frame for all movies
   - Fill the genre data frame with total rating and rating times
   - Sort the genre data frame and determine the best one
3. **Recommendation System**
   - Train a suitable algorithm
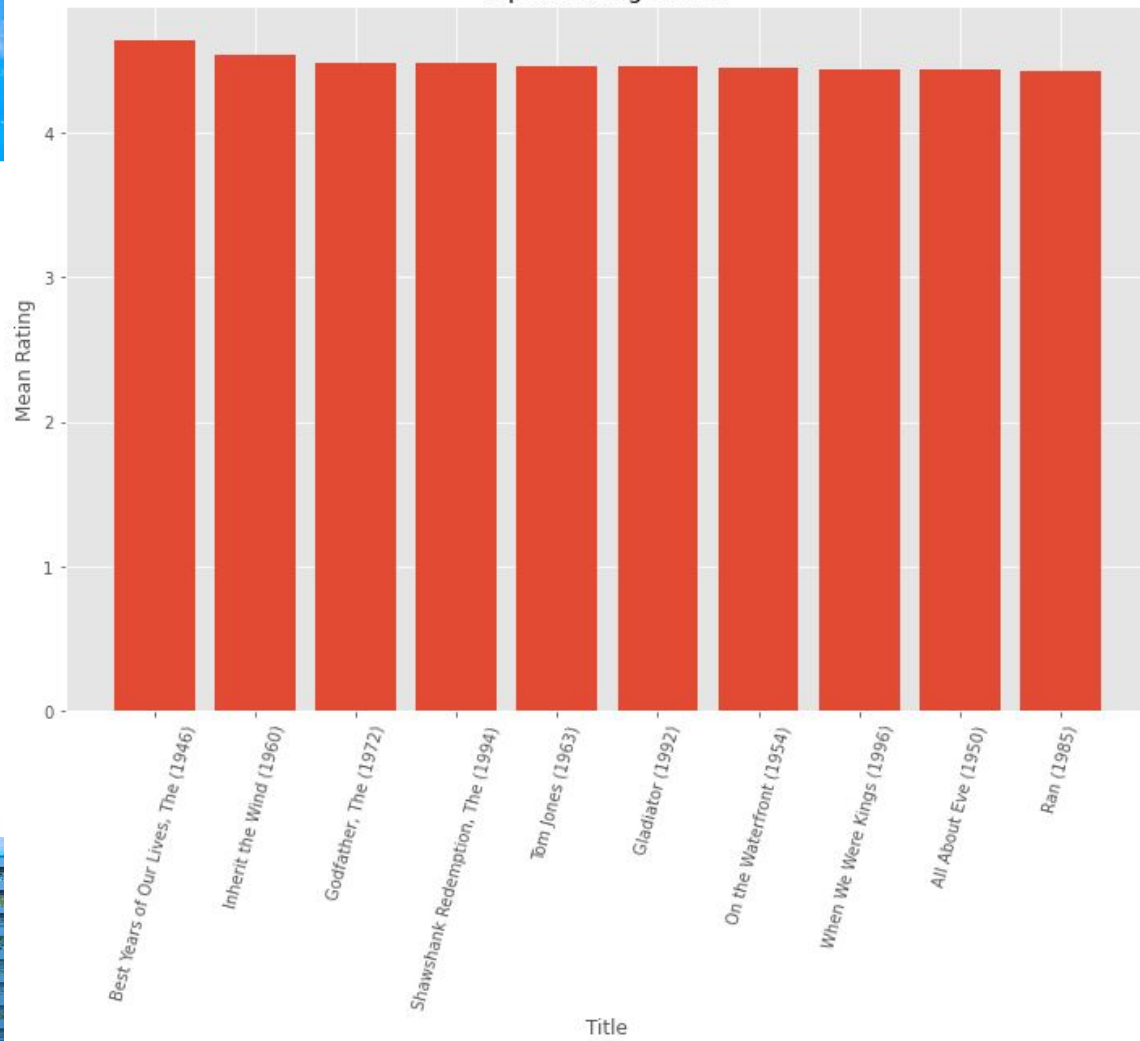   - Tune hyper parameter for this model
   - Build a recommendation system

2

1.
Best Movie

a. Movie Rating Table

b. Movie Rating Bar Graph

c. Summary

movieId	rating	rating_times	title	genres
1939	4.636364	11	Best Years of Our Lives, The (1946)	Drama|War
3469	4.541667	12	Inherit the Wind (1960)	Drama
858	4.487500	200	Godfather, The (1972)	Crime|Drama
318	4.487138	311	Shawshank Redemption, The (1994)	Crime|Drama
1948	4.458333	12	Tom Jones (1963)	Adventure|Comedy|Romance
8132	4.454545	11	Gladiator (1992)	Action|Drama
1945	4.448276	29	On the Waterfront (1954)	Crime|Drama
1147	4.437500	16	When We Were Kings (1996)	Documentary
926	4.434211	38	All About Eve (1950)	Drama
1217	4.423077	26	Ran (1985)	Drama|War

Top 10 Rating Movies

Summary

Best Movie: Godfather, The(1972) and Shawshank Redemption, The(1994)

Godfather, The(1972) :
Mean Rating: 4.487500
Rating Times: 200
Genre: Crime, Drama

Shawshank Redemption, The(1994) :
Mean Rating: 4.487138
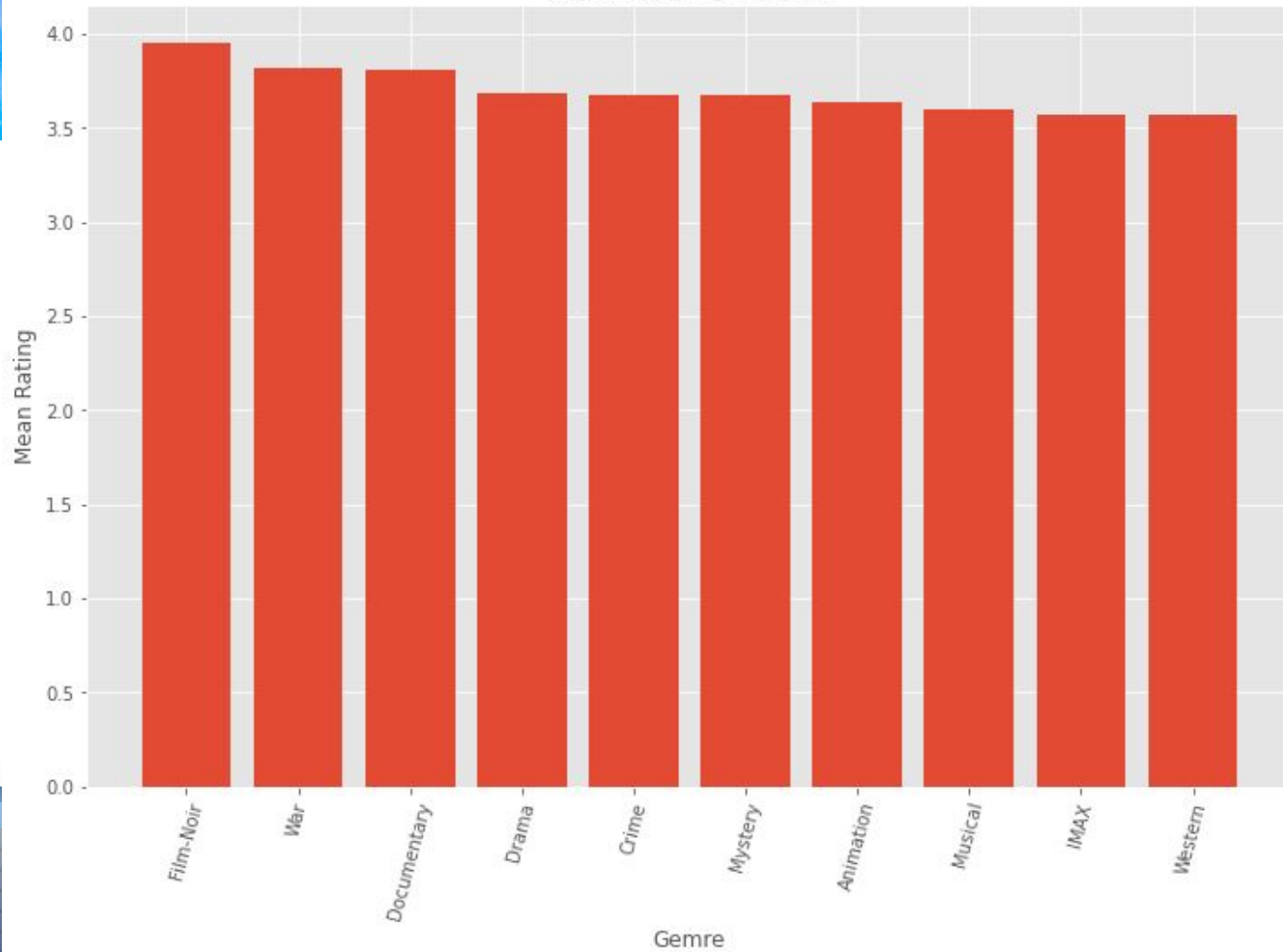Rating Times: 311
Genre: Crime, Drama

2.
Best Genre

a. Genre Rating Table

b. Genre Rating Bar Graph

c. Summary

	totalRating	meanRating	count
Film-Noir	4509.5	3.955702	1140
War	19181.5	3.817214	5025
Documentary	5964.0	3.813299	1564
(no genres listed)	68.0	3.777778	18
Drama	164767.0	3.681780	44752
Crime	59853.0	3.679639	16266
Mystery	28056.5	3.679541	7625
Animation	22434.5	3.636062	6170
Musical	16993.5	3.598793	4722
IMAX	11270.5	3.571134	3156
Western	6819.0	3.566423	1912
Romance	68762.0	3.556165	19336
Adventure	77508.5	3.520393	22017
Thriller	88807.0	3.518502	25240
Fantasy	37492.0	3.518063	10657
Children	30086.5	3.466187	8680
Sci-Fi	53169.5	3.460430	15365
Comedy	131039.0	3.446037	38026
Action	93224.5	3.445613	27056
Horror	22510.5	3.315243	6790

Top 10 Rating Genres

Summary

Best Genre: Film-Noir

Mean Rating: 3.955702
Rating Times: 1140

3.
Recommendation System 🏃

a. Algorithm Training(appendix)

b. Hyperparameter Tuning(appendix)

c. Recommendation System

```
# Obtain user ratings
user_rating = movie_rater(movies, 4, 'Comedy')
```

```
       movieId                    title              genres
8159     97328  Liberal Arts (2012)  Comedy|Drama
How do you rate this movie on a scale of 1-5, press n if you have not seen :
4
       movieId                                        title              genres
3697      4717  Battle Creek Brawl (Big Brawl, The) (1980)  Action|Comedy
How do you rate this movie on a scale of 1-5, press n if you have not seen :
5
       movieId                              title              genres
4221      5604  Man in the White Suit, The (1951)  Comedy|Sci-Fi
How do you rate this movie on a scale of 1-5, press n if you have not seen :
2
       movieId                            title              genres
5467      8596  Revenge of the Pink Panther (1978)  Comedy|Crime
How do you rate this movie on a scale of 1-5, press n if you have not seen :
4
```

```
recommended_movies(ranked_movies,movies,5)
```

```
Recommendation #  1 :   931      Cinema Paradiso (Nuovo cinema Paradiso) (1989)
Name: title, dtype: object

Recommendation #  2 :   733      Casablanca (1942)
Name: title, dtype: object

Recommendation #  3 :   984      Raging Bull (1980)
Name: title, dtype: object

Recommendation #  4 :   977      Godfather: Part II, The (1974)
Name: title, dtype: object

Recommendation #  5 :   1001      Graduate, The (1967)
Name: title, dtype: object
```

Recommendation

For Users:
* Some best movies are in crime and drama genres. And these two genres both having around 3.68 mean rating which are the top five genres in our movies data.
* If you don't have any idea what genre movies to w[...] film-noir will be your best choice because it has the highest mean rating(3.955702) with 1140 rating time[...]
* If you need more accuracy recommendation, you can use the recommendation system.
For analysts:
* SVDpp is an accuracy algorithm in this case but th[...] shortcoming is costing a long run time.
* GridsearchCV is a handy tool to help us tune mod[...] parameters, but the shortcoming also is costing a lo[...] run time.

When watching funny movie alone

When watching funny movie with family
Ha Ha Ha

Future Work

Algorithm: Try more algorithms from other packages

Hyper Parameter Tuning: Tune more parameters of the models and use random search before grid search

Exploratory Data Analysis: Exploratory more characteristics from the dataset such as timestamp column

Thank you!

Any questions?

You can find me at:

https://www.linkedin.com/in/jiarong-jr-chen-ba87b214a/

https://github.com/JRRRRRRR

Appendix

Algorithm	RMSE	MAE
KNN Basic	0.9684	0.7448
KNN Base Line	0.8958	0.6862
KNN Means	0.9186	0.7033
KNN ZScore	0.9163	0.6976
SVD	0.8957	0.6898
SVDpp	0.8884	0.6814
NMF	0.9451	0.7266
SlopeOne	0.9294	0.7117
CoClustering	0.9647	0.7478

{'rmse': 0.8854225634526053, 'mae': 0.6789764937220142}
{'rmse': {'n_epochs': 10, 'lr_all': 0.009}, 'mae': {'n_epochs': 20, 'lr_all': 0.007}}