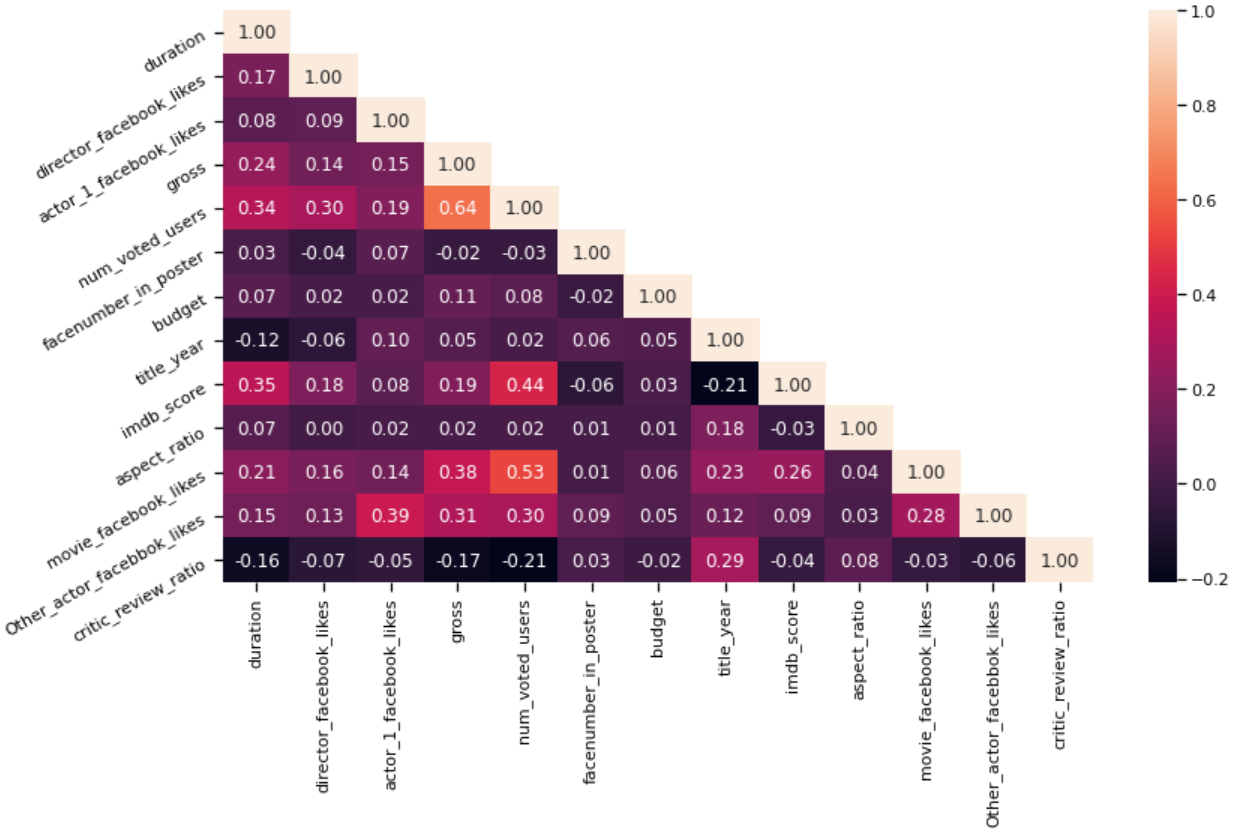
Report:



We can observe that there is a strong correlation between the actor 1 facebook like and the cast total facebook likes. Actors 2 and 3 are also connected to the sum in some way. In order to make two variables out of them, actor 1 facebook likes and other actors facebook likes, we will alter them.

Number of voted users, Number of Users for Reviews, and Number of Critics for Reviews all have strong connections. We wish to preserve the number of voters and take the proportion of reviewers to critics.

The success of these movies depends on various factors like budget, director, actor, etc. However, it has become a trend to predict the rating of the movie based on the data collected from social media

related to the movie. This will help a number of businesses relying on the movie industry in making promotional and marketing decisions.

In this report, the aim is to collect movie data from IMDB and its social media data and compare the performance of five machine learning algorithms –

1. Logistic
2. KNN
3. Decision Tree
4. Ada Boosting
5. Random Forests

Logistic Reports

	precision	recall	f1-score	support
1	0.00	0.00	0.00	52
2	0.54	0.42	0.47	387
3	0.76	0.89	0.82	921
4	0.94	0.55	0.69	62
accuracy			0.72	1422
macro avg	0.56	0.47	0.50	1422
weighted avg	0.68	0.72	0.69	1422

KNN Reports

	precision	recall	f1-score	support
1	0.00	0.00	0.00	52
2	0.51	0.45	0.48	387
3	0.73	0.85	0.79	921
4	1.00	0.24	0.39	62
accuracy			0.68	1422
macro avg	0.56	0.39	0.41	1422
weighted avg	0.66	0.68	0.66	1422

Decision Tree Reports

	precision	recall	f1-score	support
1	0.24	0.17	0.20	52
2	0.50	0.56	0.53	387
3	0.79	0.75	0.77	921
4	0.53	0.66	0.59	62
accuracy			0.67	1422
macro avg	0.52	0.54	0.52	1422
weighted avg	0.68	0.67	0.68	1422

Ada Boosting

	precision	recall	f1-score	support
1	0.26	0.17	0.21	52
2	0.51	0.56	0.53	387
3	0.79	0.75	0.77	921
4	0.56	0.71	0.62	62
accuracy			0.68	1422
macro avg	0.53	0.55	0.53	1422
weighted avg	0.68	0.68	0.68	1422

Random Forests Reports

	precision	recall	f1-score	support
1	1.00	0.02	0.04	52
2	0.66	0.50	0.57	387
3	0.78	0.93	0.85	921
4	0.91	0.52	0.66	62
accuracy			0.76	1422
macro avg	0.84	0.49	0.53	1422
weighted avg	0.76	0.76	0.73	1422

XGBoosting

	precision	recall	f1-score	support
1	0.40	0.08	0.13	52
2	0.61	0.62	0.62	387
3	0.82	0.86	0.84	921
4	0.82	0.58	0.68	62
accuracy			0.76	1422
macro avg	0.66	0.54	0.57	1422
weighted avg	0.74	0.76	0.75	1422

TESTING AND EVALUATION RESULTS To calculate the performance metrics, K-fold cross validation was performed using the algorithms on the training data that was randomly selected from datasets. The performance score was averaged over all the scores for each model and datasets.

Random Forest and XGBoosting was successful in predicting reviews and had excel lent F1 scores. Ada Boosting had 0.68 F1 scores for same dataset. Here, the F1 acc uracy score of Random Forest and XGBoosting was 0.76 Thus Random Forest is th e best option for Prediction .

Algorithms	Accuracy
Logistic	0.7165963431786216
KNN	0.6828410689170182
Decision Tree	0.6744022503516175
XGBoosting	0.759493670886076
Random Forests	0.7559774964838256