

1. import packages (numPy, pandas, matplotlib, seaborn)
 2. Load 'movie_metadata.csv'
 3. Drop irrelevant columns
 4. Deal with missing values
 5. Remove the duplicate value
 6. Reorganize several columns
 7. Deal with correlation problem
 8. Deal with categorical columns (dummy)
 9. Data virtualization
 10. split dataset into training and testing dataset
 11. Create models (Logistic Regression, KNN, Random Forest)
 12. Model comparison
 13. Conclusion
- (#The conclusion is that Random Forest Algorithm have best accuracy which is around 77%)