MOVIE RECOMMENDATIONS

GROUP 5: TAHA MIRGHORBANI,
HAGAN PRESTINARIO,
AND NOVA MITCHELL

PROBLEM

PREDICTING POTENTIAL MOVIE RECOMMENDATIONS BASED ON THE USER'S POTENTIAL MOVIE RATING

Dataset

Target Variable: Vote Rating

- High_rated (Avg_vote > 7)
- Avg_rated (7 > avg_vote > 5)
- Low_rated (Avg_vote> 5)

Features:

- -Review from useers, critics, and overall average vote
- -Boxoffice and related financial performance
- -Genre
- -Region, and language

DATA SPECIFICATIONS

7250 movies

46 Record

DATA CLEANING

Read Data

1
Filter
Filter data to keep the budget values that are US dollar format

2
Drop
Drop inapplicable columns

3
Drop
Drop missing values

4
Remove
Remove the dollar sign from the 'budget', 'usa_gross_income', and 'worlwide_gross_income' columns

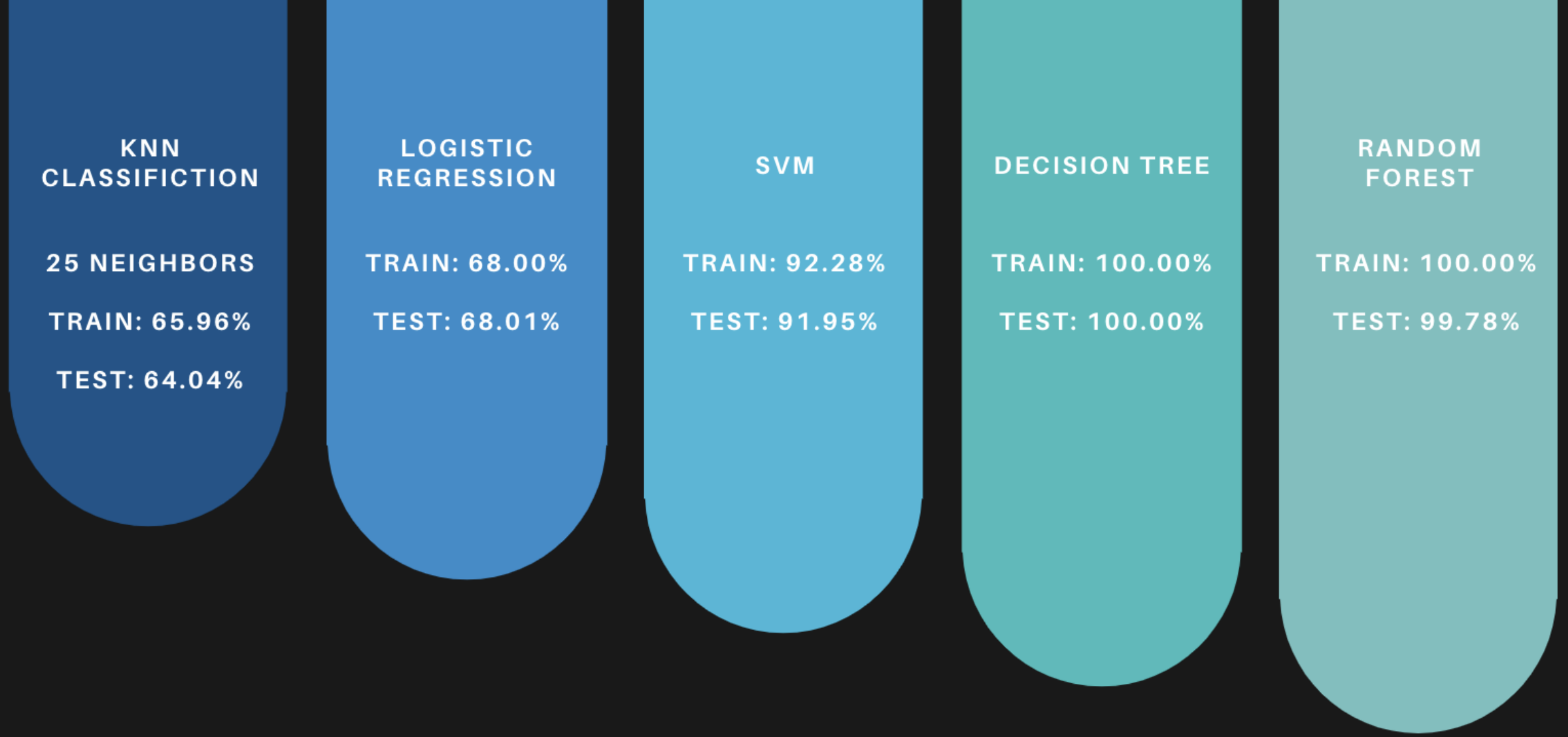
5
Split/Categorize
Split up the listed movie genres, remove duplicates, and categorize

6
Specify/Categorize
Specify the countries that will be used and categorize

7
Specify/Categorize
Specify the languages that will be used and categorize

8
Categorize
Categorize the average vote value as: high_rated, avg_rated, low_rated

Create Features and Target Sets



PREDICTIVE MODEL - RESULTS

FEATURE IMPORTANCE

DECISION TREE

FEATURES	IMPORTANCE
AVG_VOTE	1
DURATION	0
MEXICO	0
FILM-NOIR	0
COMEDY	0
MUSICAL	0
USA	0
FRANCE	0
JAPAN	0
UK	0
ITALY	0
AUSTRALIA	0
CANADA	0
SPAIN	0
GERMANY	0
RUSSIA	0
ENGLISH	0
SPANISH	0
FRENCH	0
GERMAN	0
ITALIAN	0
RUSSIAN	0
HISTORY	0
ACTION	0
ROMANCE	0
MYSTERY	0
VOTES	0
BUDGET	0
USA_GROSS_INCOME	0
WORLWIDE_GROSS_INCOME	0
REVIEWS_FROM_USERS	0
REVIEWS_FROM_CRITICS	0
HORROR	0
THRILLER	0
FAMILY	0
ANIMATION	0
CRIME	0
BIOGRAPHY	0
SCI-FI	0
FANTASY	0
DRAMA	0
ADVENTURE	0
WESTERN	0
WAR	0
SPORT	0
JAPANESE	0

RANDOM FOREST

FEATURES	IMPORTANCE
AVG_VOTE	0.663589
VOTES	0.065735
REVIEWS_FROM_USERS	0.043055
DURATION	0.033152
REVIEWS_FROM_CRITICS	0.02973
BUDGET	0.02781
USA_GROSS_INCOME	0.026545
WORLWIDE_GROSS_INCOME	0.024363
DRAMA	0.014639
HORROR	0.00546
BIOGRAPHY	0.005285
ACTION	0.004957
COMEDY	0.003811
ENGLISH	0.003577
USA	0.003508
ADVENTURE	0.003033
THRILLER	0.002656
CRIME	0.002626
UK	0.002482
ROMANCE	0.002306
FANTASY	0.002274
FAMILY	0.002228
ANIMATION	0.002143
CANADA	0.001977
SPANISH	0.001908
MYSTERY	0.001756
FRENCH	0.001715
SCI-FI	0.001679
FRANCE	0.00162
GERMANY	0.001516
GERMAN	0.001468
ITALIAN	0.001258
JAPAN	0.001179
RUSSIAN	0.001053
ITALY	0.000991
JAPANESE	0.000979
HISTORY	0.000891
MUSICAL	0.000876
WAR	0.00081
MEXICO	0.000758
SPORT	0.000739
AUSTRALIA	0.000737
SPAIN	0.000498
RUSSIA	0.000351
WESTERN	0.00027
FILM-NOIR	0.000006

PREDICTION FOR TARGET VARIABLE

	Toy Story	Greta	Street Fighter
KNN	High	Average	Average
Logistic	High	Average	Average
SVM	High	Average	Low
Data Tree	High	Average	Low
RandomForest	High	Average	Low

RECOMMENDATION

Yes!

+ SVM
+ Random Forest
+ Data Tree

No

+ Logistic Regression
+ KNN

THANK YOU

We will now answer any questions