Game Application Rating Prediction

Artificial Intelligence

CS-617-C

GILLELLA AVINASH REDDY

Submitted To: Dr. Reza Sadeghi

Spring 2022

Introduction

This projects aims predict the rating of game applications in an application store that can be achieved in course of time. Here we consider main attributes like customer rating, customer count, in app purchases so on which supports our prediction. This helps in us to find out the best game genre to be released in a certain area to get a best customer rating.

Research Question

How gaming companies can attain market attention?

How to predict rank of game application accessible through App Store?

DATA DESCRIPTION

Data set contains 1707 samples and 17 features.

Source of data:
[https://verzeo.com/]

Exploring data

It refers to the critical process of performing initial investigations on data so as to discover patterns,to spot anomalies,to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

head( )

```
In [4]: df.head(3)
```

Out[4]:

	Icon URL	Average User Rating	User Rating Count	Price	In-app Purchases	Description	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
	https://is2-tic.com/image/thumb/Purpl...	4.0	3553.0	2.99	NaN	Join over 21,000,000 of our fans and download ...	Mighty Mighty Good Games	4+	DA, NL, EN, FI, FR, DE, IT, JA, KO, NB, PL, PT...	15853568.0	Games	Games, Strategy, Puzzle	11/07/2008	30/05/2017
	https://is4-tic.com/image/thumb/Purpl...	3.5	284.0	1.99	NaN	The classic game of Reversi, also known as Oth...	Kiss The Machine	4+	EN	12328960.0	Games	Games, Strategy, Board	11/07/2008	17/05/2018
	https://is5-tic.com/image/thumb/Purpl...	3.0	8376.0	0.00	NaN	Play the classic strategy game Othello (also k...	Bayou Games	4+	EN	674816.0	Games	Games, Board, Strategy	11/07/2008	5/09/2017

DIsplays the top 3 samples of the data set

describe( )

```
In [30]: df.describe()
```

Out[30]:

	Average User Rating	User Rating Count	Price	Size	Primary Genre	Description_word_length	various_size
count	7488.000000	7.488000e+03	7488.000000	7488.000000	7488.000000	7488.000000	7488.000000
mean	4.062099	3.306245e+03	0.569686	144.545651	6.091079	2.366854	2.929754
std	0.750506	4.251578e+04	2.422359	244.092470	1.455905	0.782749	0.265321
min	1.000000	5.000000e+00	0.000000	0.205841	0.000000	1.000000	1.000000
25%	3.500000	1.200000e+01	0.000000	29.066406	6.000000	2.000000	3.000000
50%	4.500000	4.600000e+01	0.000000	75.625000	6.000000	3.000000	3.000000
75%	4.500000	3.072500e+02	0.000000	169.050049	6.000000	3.000000	3.000000
max	5.000000	3.032734e+06	139.990000	3820.029297	20.000000	3.000000	3.000000

The Describe function **returns the statistical summary of the dataframe or series**. This includes count, mean, median (or 50th percentile) standard variation, min-max, and percentile values of columns

Info( )

```
In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17007 entries, 0 to 17006
Data columns (total 18 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   URL                           17007 non-null  object
 1   ID                            17007 non-null  int64
 2   Name                          17007 non-null  object
 3   Subtitle                      5261 non-null   object
 4   Icon URL                      17007 non-null  object
 5   Average User Rating           7561 non-null   float64
 6   User Rating Count             7561 non-null   float64
 7   Price                         16983 non-null  float64
 8   In-app Purchases              7683 non-null   object
 9   Description                   17007 non-null  object
 10  Developer                     17007 non-null  object
 11  Age Rating                    17007 non-null  object
 12  Languages                     16947 non-null  object
 13  Size                          17006 non-null  float64
 14  Primary Genre                 17007 non-null  object
 15  Genres                        17007 non-null  object
 16  Original Release Date         17007 non-null  object
 17  Current Version Release Date  17007 non-null  object
dtypes: float64(4), int64(1), object(13)
memory usage: 2.3+ MB
```

The info() method **prints information about the DataFrame**. The information contains the number of columns, column labels, column data types, memory usage, range index, and the number of cells in each column (non-null values).

shape( )

```
In [7]: df.shape
Out[7]: (17007, 18)
```

Describes the number of samples and features in the data frame

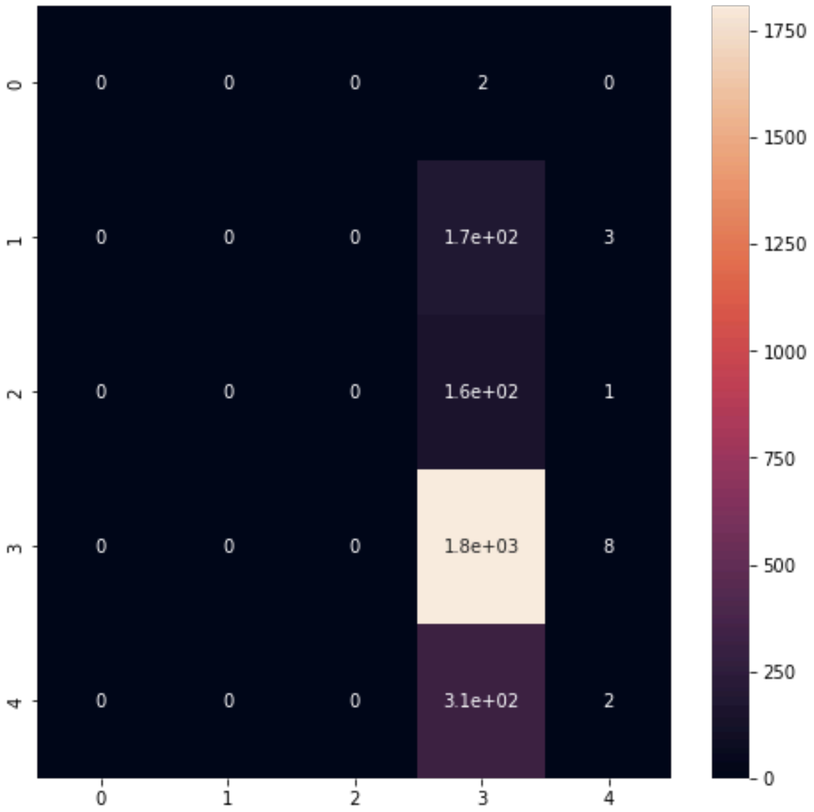
Modeling data

In this project, I have developed 3 model using

- K Nearest Neighbour Classification
- Support Vector Classification
- Logistic Regression

- K Nearest Neighbour Classifier:

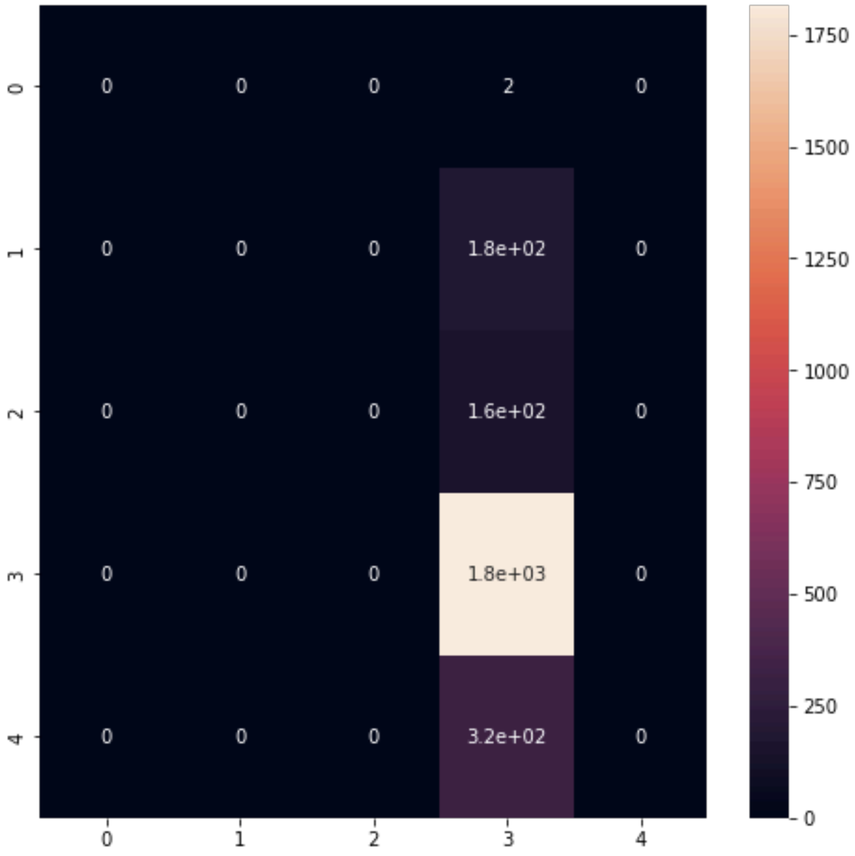
  Accuracy - 73.2%



confusion matrix

- Support Vector Classifier:

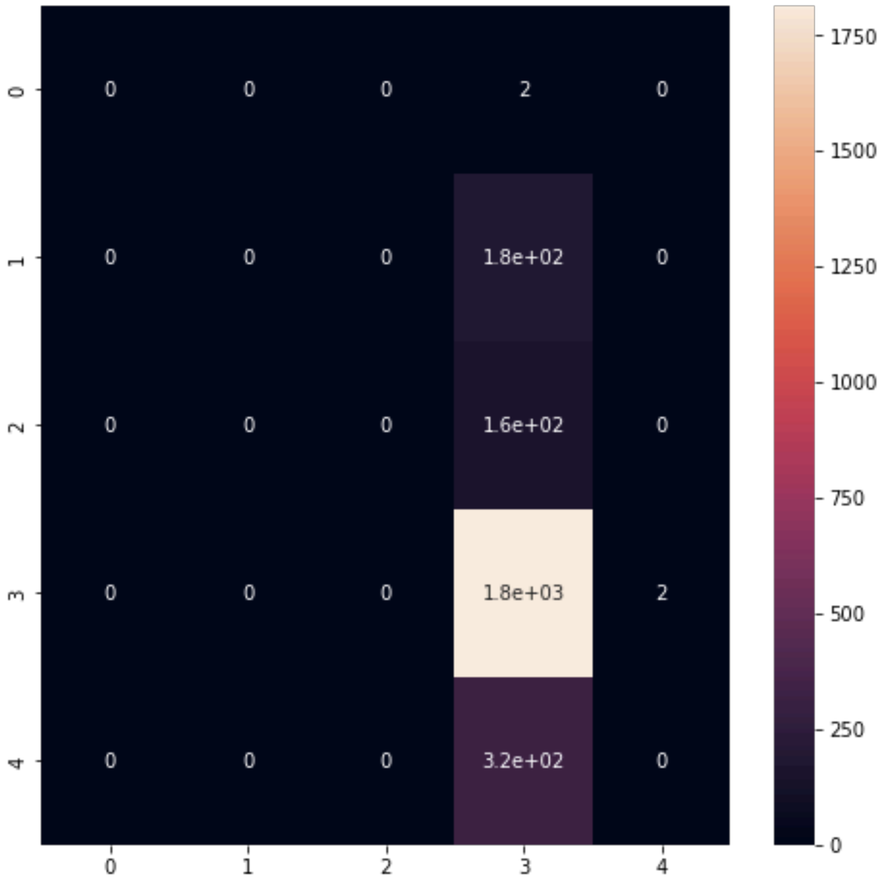
  Accuracy - 73.46%



confusion matrix

- Logistic Regression :

  Accuracy - 73.38%



confusion matrix

Optimisation

- K-Folds cross-validator
- Creating new features
- Removing outliers

CONCLUSION

To conclude that most of the users are interested in game applications which are either free or have low price. Rating ranges from 4 to 5 for most of the games with average application size.

GitHub Repository Address

https://github.com/avinashgillella/avinashgillella-Game-Application-rating-prediction.git

Thank you