

Final Project Analysis on Metacritic games

The motivation behind looking at this data set is to see if there is a trend in how games are being made. To do so our group will look at the variables within the data and apply machine learning techniques learned throughout the course to see if any connections and predictability can be made.

The problem at hand is that certain game genres are seeing a recent oversaturation within the game market. This game genre is called Battle Royale. Some recent titles in this particular genre would be "Player Unknown's Battlegrounds" and "Apex Legends." These are two of the front runners in this relatively new and upcoming genre.

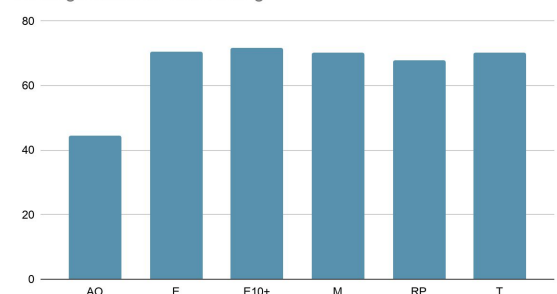
Our proposed method is to see whether or not there is a predictability of whether certain types of games do well within critic and player reviews based on the variables given to us in the data set. To do this we will use the CART machine learning algorithm to see if there is predictability within said variables.

We used machine learning through the cart method using the 'Rating' variable as the independent variable. Rating is based off of the game rating being: E - for everyone, T - for teen, M - for mature, etc. We then used the critic review columns positive, neutral, and negative as well as the user review columns positive, neutral, and negative along with the metascore column and user score column as the dependent variables. When we entered said variables into machine learning using the CART method, we got back an accuracy Train score of 0.9994(rounded to 4 decimal places). Our first test result came back with an accuracy rating of 0.3177(rounded to 4 decimal places). This is a low accuracy score and as a result of that we tried dropping different columns from the independent variable. Our highest accuracy score we achieved was 0.3348(rounded to 4 decimal places). after dropping both the 'neutral critics' and 'neutral user' scores. This means that machine learning is not a great way to predict rating results in better reviews from both critics and players.

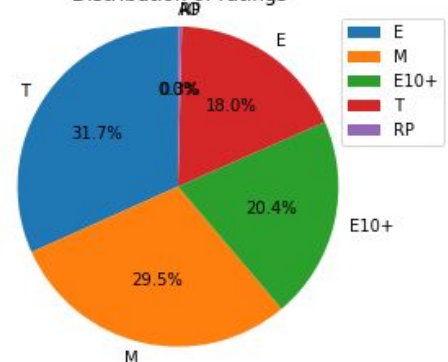
confusion_matrix if

[39	24	33	1	32]
[29	37	30	3	44]
[35	37	79	1	52]
[1	0	0	0	0]
[42	48	63	1	74]]

Average score for each rating

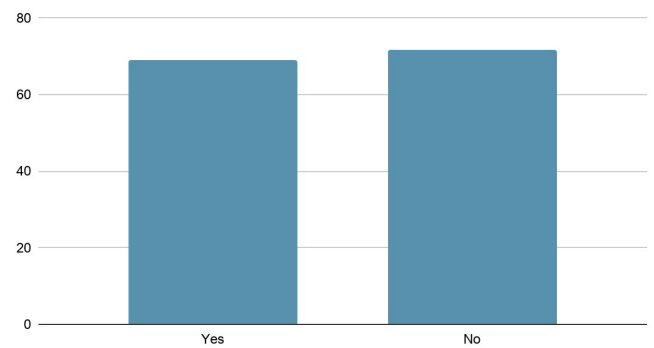


Distribution of ratings



We used machine learning through the cart method using the 'number players' variable as the independent variable. 'Number players' is determined on the game online multiplayer or a range of co-op ranging from 1-10 players in the same game from the same console. We then used the critic review columns positive, neutral, and negative as well as the user review columns positive, neutral, and negative along with the metascore column and user score column as the dependent variables. When we entered said variables into machine learning using the CART method, we got back an accuracy Train score of 1. Our first test result came back with an accuracy rating of 0.5475(rounded to 4 decimal places). This is a relatively accurate accuracy score and we determined it would be acceptable for predicting using machine learning. This means that machine learning is a useful way to predict if a game has multiplayer or not results in better reviews from both critics and players.

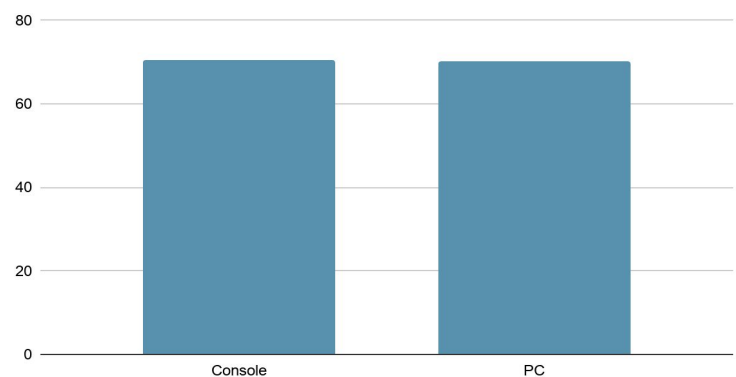
Multiplayer vs Gamescore



```
confusion_matrix
[[231 177]
 [142 155]]
```

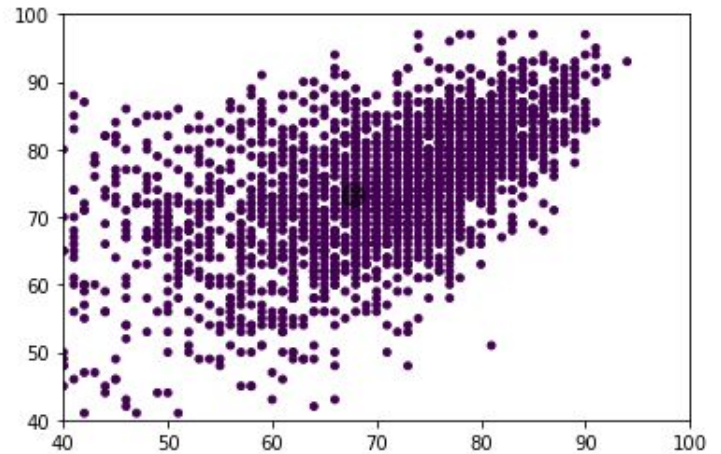
We used machine learning through the CART method using the 'platform' variable as the independent variable. Platform is based off of the game rating being: PC, PS4, Xbox One, Nintendo Switch, etc. We then used the critic review columns positive, neutral, and negative as well as the user review columns positive, neutral, and negative along with the metascore column and user score column. When we entered said variables into machine learning using the CART method, we got back an accuracy Train score of 1. The accuracy score of our test data, however, gave us an accuracy score of 0.7163 (rounded to 4 decimal places). This is a high accuracy score and means that machine learning is a great way to predict if the platform results in better reviews from both critics and players.

Console gamescore vs PC gamescore



```
confusion_matrix
[[448  99]
 [101  57]]
```

In conclusion we found that only in two of the three variables given in the dataset were we able to find moderate predictability. One of the three variables being 'rating' we found very low predictability being '.3177'. This shows that the game rating is not a huge factor on how well the game is received by both players and critics alike. The second variable we looked at in terms of predictability was the platform the game was released on. We found that games were reviewed slightly higher on consoles over other platforms being PC. The accuracy score we got back after applying to machine learning was '0.7163' proving that it is fairly predictable when releasing games for differing platforms. Lastly we compared how games were reviewed based on what style of multiplayer they had, if any. With this as the independent variable we found an accuracy score of '0.5475' which shows that this variable has some impact on what sort of reviews it gets from both critics and players alike.



Our KMeans plot we made came out to be all clustered in one giant cluster. Along with this the scatter plot points seem to be in a grid format, this is because the review variables (positive, neutral, and negative) are all int values with no floating point numbers. This shows in the scatter plot by the points not varying from integer numbers, resulting in a grid format.

- Each of us did about half of the actual code
- Carter Meseck made the presentation
- Matt Moran wrote the Final Project