Machine Learning

Predicting the numerical ratings for Animal Crossing reviews based on their text

How will we be feeling today?

- Predict the rating or grade of a review based on keywords.
- It is difficult to sort through and categorize thousands of user and critic reviews manually, so accurate machine learning is helpful in automating this tedious task.
- This has applications in analyzing data without numerical ratings, such as social media posts and comments that only use text to grade the game.
- Follow along with the Jupyter notebook!

Patterns and Prediction

```
critics.describe
```

	grade
count	107.000000
unique	NaN
top	NaN
freq	NaN
mean	90.635514
std	6.114308
min	70.000000
25%	90.000000
50%	90.000000
75%	94.000000
max	100.000000

```
users.describe(ir
```

	grade
count	2999.000000
unique	NaN
top	NaN
freq	NaN
mean	4.217406
std	4.349486
min	0.000000
25%	0.000000
50%	2.000000
75%	10.000000
max	10.000000

	grade	publication	
0	100	Pocket Gamer UK	Animal Crossing; New Horizons, much
1	100	Forbes	Know that if you’re overwhelmed with
2	100	Telegraph	With a game this broad and lengthy,
3	100	VG247	Animal Crossing: New Horizons is eve
4	100	Nintendo Insider	Above all else, Animal Crossing: New
5	100	Trusted Reviews	Animal Crossing: New Horizons is the
6	100	VGC	Nintendo's comforting life sim is a
7	100	God is a Geek	A beautiful, welcoming game that is
8	100	Nintendo Life	Animal Crossing: New Horizons takes
9	100	Daily Star	Similar to how Breath of the Wild an

	grade	user_name	
0	4	mds27272	My gf started playing before me. No option
1	5	lolo2178	While the game itself is great, really rel
2	0	Roachant	My wife and I were looking forward to play
3	0	Houndf	We need equal values and opportunities for
4	0	ProfessorFox	BEWARE! If you have multiple people in yo
5	0	tb726	The limitation of one island per Switch (r
6	0	Outryder86	I was very excited for this new installmen
7	0	Subby89	It's 2020 and for some reason Nintendo has
8	0	RocketRon	This is so annoying. Only one player has t
9	0	chankills	I purchased this game for my household (me

Patterns and Prediction

- Some type of standardization or normalization for user and critic grade columns will be needed, as the critics rate on a scale of 0-100 while the users rate on a scale of 0-10. This affects differences in the mean, standard deviation, and max grade value.
- As there is a lot of textual data, I will have challenges with finding tools and setting up the ML algorithm to parse words for identifying sentiment.

Reading A Crowded Room

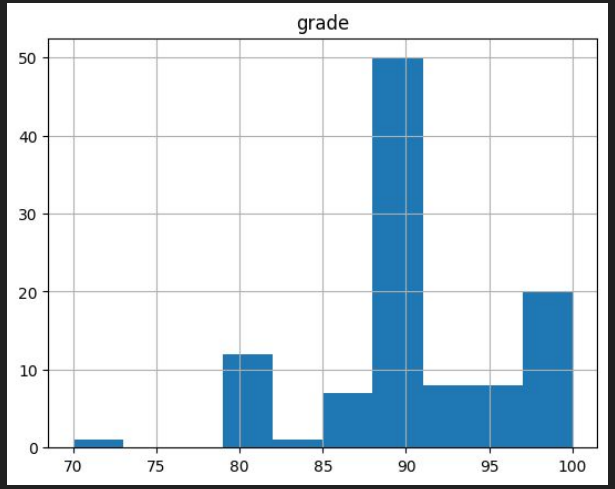
- Not many emotional words show up in either critic or user review word clouds, so additional word sorting and identification may be needed.
- User reviews do include "fun", "love", "good", bad", and "ridiculous" among most used words. The critic reviews use "love", "fun", "fantastic", "improve", and "charm" a lot.
- What features of the game might users and critics be talking about the most?

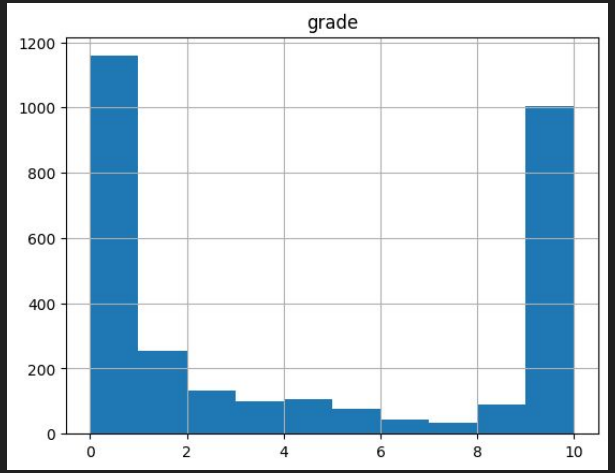
Differences between Users and Critics

- The majority of critics give high scores in the 80s and 90s, whereas the majority of users give extremely low scores (0) or extremely high scores (10).

- I wonder if ML can give insight into why this is so through word detection and the mapping of words to positive or negative sentiment.

- What subjects do critics and users have differing feelings on, and therefore what do they prioritize differently in Animal Crossing?

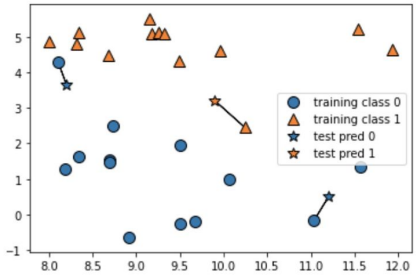




Building the Model

- This can be seen as an unsupervised learning problem, since the reviews don't already have positive or negative labels.
- This can be worked with an ML model that can infer this based on the given scores.
- However, both unsupervised and supervised learning algorithms can be used for this project, either simultaneously or choosing to do one or the other.
- These include clustering, decision trees, k-fold cross validation…

Clustering



In clustering, we don't have the colors (labels).

Rather, the algorithm comes up with the color and shape of each point.

Supervised learning.