



Machine Learning In Action

UCSD DV/DA Bootcamp 2021 Summer:
Anny Tritchler, Melissa Cardenas, Ryan Mangeno

Our conundrum...

Many board game enthusiasts spend \$50+ per game... just to find out they don't enjoy it.



Can machine learning help?



It's been done before...

Collaborative recommendation systems are currently used by companies such as:

- Google
- Yelp
- Netflix
- Zillow

So.. this is totally possible, right?

The Data

From Kaggle: The 20,000 Boardgames Dataset

Data originally scraped from boardgamegeek.com

objectid	name	average	avgweight	boardgamecategory	boardgamemechanic
0 174430	Gloomhaven	8.85292	3.8078	['Adventure', 'Exploration', 'Fantasy', 'Fight...	['Campaign / Battle Card Driven', 'Cooperative...
1 161936	Pandemic Legacy Season 1	8.62499	2.8301	['Environmental', 'Medical']	['Action Points', 'Cooperative Game', 'Hand Ma...
2 167791	Terraforming Mars	8.42299	3.2313	['Economic', 'Environmental', 'Industry / Manu...	['Card Drafting', 'End Game Bonuses', 'Hand Ma...
3 182028	Through the Ages A New Story of Civilization	8.49419	4.3850	['Card Game', 'Civilization', 'Economic']	['Action Points', 'Auction/Bidding', 'Auction:...
4 224517	Brass Birmingham	8.62031	3.9122	['Economic', 'Industry / Manufacturing', 'Tran...	['Hand Management', 'Income', 'Loans', 'Market...
...
19995 5154	Franchise	6.00000	0.0000	['Economic', 'Negotiation']	['Simulation', 'Trading']
19996 5158	Punto y Raya	4.00000	0.0000	['Abstract Strategy']	['Enclosure']
19997 5159	3 Up	6.00000	0.0000	['Abstract Strategy', 'Childrens Game']	['Pattern Building']
19998 5160	Dino-Checkers	5.77500	0.0000	['Abstract Strategy', 'Childrens Game']	['Area Majority / Influence']
19999 5163	Obsession	4.60000	2.0000	['Abstract Strategy']	['Dice Rolling', 'Pattern Building']
20000 5164

The Problems... .so many problems..

Unicode errors

3076	Puerto Rico	8.02082	3.283
7736	A Feast for Odin	8.18223	3.846
6848	Mage Knight Board Game	8.10013	4.284
4928	Orléans	8.08174	3.063
2794	Caverna The Cave Farmers	8.04560	3.781
6192	Wingspan	8.10343	2.365
5014	Food Chain Magnate	8.17751	4.197

Lists that aren't lists

boardgamecategory	boardgamemechanic
['Adventure', 'Exploration', 'Fantasy', 'Fight... [Environmental', 'Medical']	['Campaign / Battle Card Driven', 'Cooperative... ['Action Points', 'Cooperative Game', 'Hand Ma...
conomic', 'Environmental', 'Industry / Manu... ['Card Game', 'Civilization', 'Economic']	['Card Drafting', 'End Game Bonuses', 'Hand Ma... ['Action Points', 'Auction/Bidding', 'Auction:...
conomic', 'Industry / Manufacturing', 'Tran...	['Hand Management', 'Income', 'Loans', 'Market...

Looks like lists, right? Nope - they're strings!

Data Preparation

➤ Get_dummies

Category_'Abstract Strategy'	Category_'Action / Dexterity'	Category_'Adventure'	Category_'Age of Reason'	Category_'American Civil War'	Category_'American Revolutionary War'	Category_'American West'	Category_'American West'
0	0	0	0	0	0	0	0

- Remove outliers to avoid skewed data
- Filter out and drop specific terms from each column based on value counts

```
pd.set_option('display.max_rows', 500)

cat_counts = categories_df.explode('Category')
cat_counts.columns.str.replace("'", "")
cat_vc = cat_counts.groupby("Category").size().sort_values(ascending=False)

categories_to_replace = cat_vc[cat_vc < 50].index

# # Replace in dataframe
for cats in categories_to_replace:
```

Machine Learning Results

```
knn.score(X_train, y_train)
```

✓ 3.9s

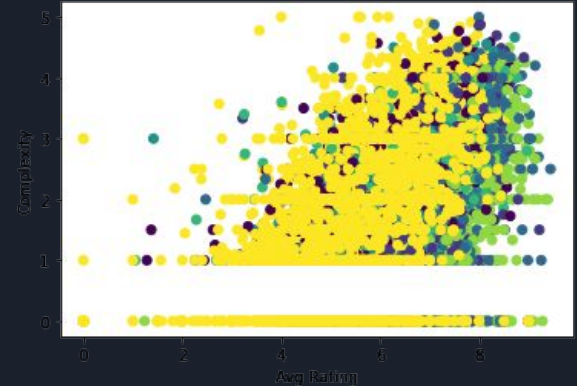
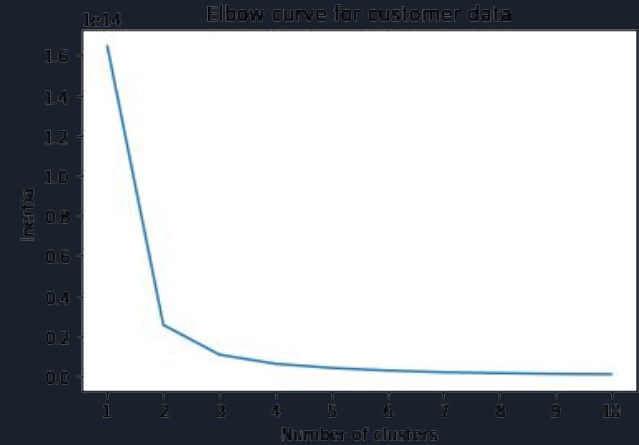
0.1702


```
model_loss, model_accuracy = nn_model.evaluate(x_train,y_train,verbose=1)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")
```

✓ 1.8s

1875/1875 [=====] - 2s 913us/step - loss: 0.2423 - accuracy: 0.9327

Loss: 0.2422737032175064, Accuracy: 0.9326500296592712





But did it really
work?

- High accuracy was only achieved when clustering on rating vs complexity. Categorical terms were not weighted.
- There's just not enough user input to properly train a machine algorithm

So what about the problem at hand?

We solved it... but without ML:

- Uses ML algorithms but not true machine learning (no training) for a cold-start recommender

```
get_recommendations('Gloomhaven', cosine_sim)
```

✓ 0.4s


11211		Frosthaven
1336	Middara	Unintentional Malum - Act 1
4453		Star Saga
12281		Tower of the Wizard King
95	Descent	Journeys in the Dark (Second Edition)
651	Dungeons & Dragons	Wrath of Ashardalon Board ...
758		Massive Darkness
2119		Dungeon Alliance
2592	Shadows of Brimstone	Forbidden Fortress
3693		Champions of Hara

Name: Name, dtype: object

Lessons Learned and Looking Forward

- Machine learning can't solve all problems... despite the hype.
- ML can be used for a recommender system but must be in collaborative or hybrid model.



A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

Thanks for joining us!

Any questions?