Amazon Gift Card Reviews

The Trials of fire!
Cole Baugh & Brey Rivera

Our Goal

Where we got the data

http://deepyeti.ucsd.edu/jianmo/amazon/index.html
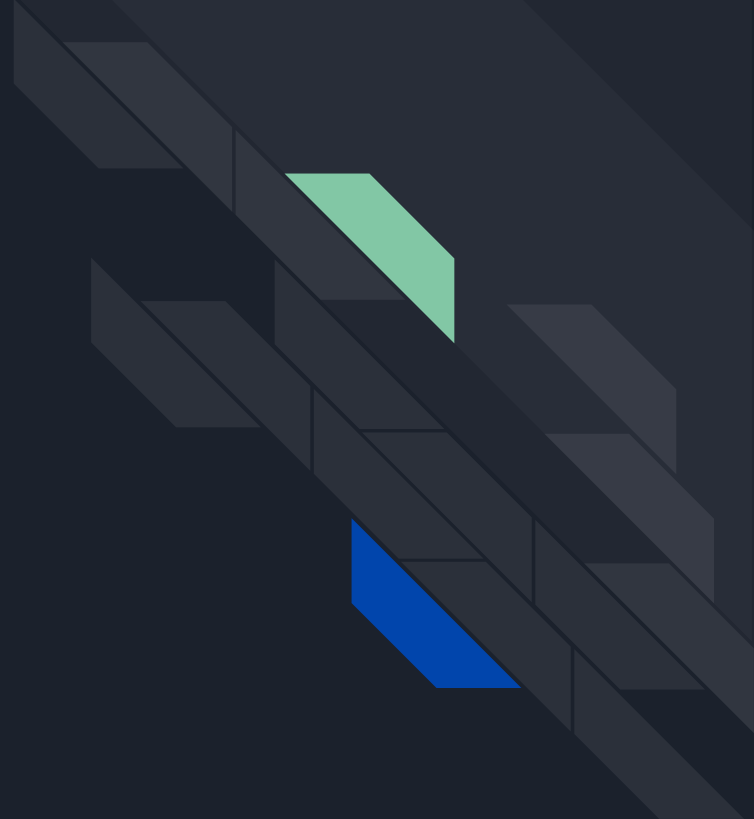
Files

Complete review data

You can download the complete dataset through the link on this website.

raw review data (34gb) - all 233.1 million reviews

ratings only (6.7gb) - same as above, in csv form without reviews or metadata

5-core (14.3gb) - subset of the data in which all users and items have at least 5 reviews (75.26 million reviews)

meta data (12gb) - meta data for all products

- We also provide a colab notebook that helps you parse and clean the data.

duplicated product list (126mb) - list of duplicated products

Per-category data - the review and product metadata for each category.

| Category	reviews	metadata
AMAZON FASHION	reviews (883,636 reviews)	metadata (186,637 products)
All Beauty	reviews (371,345 reviews)	metadata (32,992 products)
Appliances	reviews (602,777 reviews)	metadata (30,459 products)
Arts Crafts and Sewing	reviews (2,875,917 reviews)	metadata (303,426 products)
Automotive	reviews (7,990,166 reviews)	metadata (932,019 products)
Books	reviews (51,311,621 reviews)	metadata (2,935,525 products)
CDs and Vinyl	reviews (4,543,369 reviews)	metadata (544,442 products)
Cell Phones and Accessories	reviews (10,063,255 reviews)	metadata (590,269 products)
Clothing Shoes and Jewelry	reviews (32,292,099 reviews)	metadata (2,685,059 products)
Digital Music	reviews (1,584,082 reviews)	metadata (465,392 products)
Electronics	reviews (20,994,353 reviews)	metadata (786,868 products)
Gift Cards	reviews (147,194 reviews)	metadata (1,548 products)
Grocery and Gourmet Food	reviews (5,074,160 reviews)	metadata (287,209 products)
Home and Kitchen	reviews (21,928,568 reviews)	metadata (1,301,225 products)
Industrial and Scientific	reviews (1,758,333 reviews)	metadata (167,524 products)
Kindle Store	reviews (5,722,988 reviews)	metadata (493,859 products)
Luxury Beauty	reviews (574,628 reviews)	metadata (12,308 products)
Magazine Subscriptions	reviews (89,689 reviews)	metadata (3,493 products)
Movies and TV	reviews (8,765,568 reviews)	metadata (203,970 products)
Musical Instruments	reviews (1,512,530 reviews)	metadata (120,400 products)
Office Products	reviews (5,581,313 reviews)	metadata (315,644 products)
Patio Lawn and Garden	reviews (5,236,058 reviews)	metadata (279,697 products)
Pet Supplies	reviews (6,542,483 reviews)	metadata (206,141 products)
Prime Pantry	reviews (471,614 reviews)	metadata (10,815 products)
Software	reviews (459,436 reviews)	metadata (26,815 products)
Sports and Outdoors	reviews (12,980,837 reviews)	metadata (962,876 products)
Tools and Home Improvement	reviews (9,015,203 reviews)	metadata (571,982 products)
Toys and Games	reviews (8,201,231 reviews)	metadata (634,414 products)
Video Games	reviews (2,565,349 reviews)	metadata (84,893 products)

What the data looks like

	A	B	C	D	E	F	G	H	I	J	K	L
1	overall	vote	verified	reviewTime	reviewerID	asin	style	reviewerName	reviewText	summary	unixReviewTir	image
2	1	25	FALSE	12 19, 2008	APV13CM0919JD	B001GXRQW0	{'Gift Amount:': ' 50'}	LEH	Amazon,	Merry Christmas.	1229644800	
3	5		FALSE	12 17, 2008	A3G8U1G1V082SN	B001GXRQW0	{'Gift Amount:': ' 50'}	Tali	I got this gift card fr	Gift card with bes	1229472000	
4	5	4	FALSE	12 17, 2008	A11T2Q0EVTUWP	B001GXRQW0	{'Gift Amount:': ' 50'}	Z	aren't we going to	A convenient and	1229472000	
5	5		FALSE	12 17, 2008	A9YKGBH3SV22C	B001GXRQW0	{'Gift Amount:': ' 25'}	Giotravels	You can always get	Totally make sens	1229472000	
6	1		TRUE	12 17, 2008	A34WZIHVF3OKOL	B001GXRQW0		King Dad	Why take 50	Give CASH!	1229472000	
7	3	146	FALSE	12 16, 2008	A221J8EC5HNPY6	B001GXRQW0	{'Gift Amount:': ' 25'}	D. Daniels	Gift Cards are the t	Great Gift but on	1229385600	
8	1	577	TRUE	12 14, 2008	A204VFHD6JWASO	B001GXRQW0	{'Gift Amount:': ' 50'}	C. Michael Gamble	I am a dedicated Ar	Prime Members I	1229212800	
9	1	452	TRUE	12 11, 2008	A3QN3GMNS7NBIQ	B001GXRQW0	{'Gift Amount:': ' 50'}	MjSher	Bought three $50	Horrible experier	1228953600	
10	5	3	FALSE	12 10, 2008	A1BXZIDZOMBAV2	B001GXRQW0	{'Gift Amount:': ' 50'}	D. Magilnick	At this time of year	These are great g	1228867200	
11	1	45	TRUE	12 6, 2008	AJ7D8N3Z1Q7MK	B001GXRQW0	{'Gift Amount:': ' 50'}	P. Chung	I bought 8 gift	They are sealed ii	1228521600	

What the data looks like preprocessed

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	stars	image	reviewLen	disappoint	great	amazing	like	bad	waste	bought	fraud	expired	nice	easy	perfect	thank	dont	very	never
2	1	FALSE	74	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
3	5	FALSE	23	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	5	FALSE	116	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
5	5	FALSE	40	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
6	1	FALSE	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	3	FALSE	95	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
8	1	FALSE	53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
9	1	FALSE	91	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
10	5	FALSE	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	FALSE	119	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
12	5	FALSE	10	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	5	FALSE	5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
14	4	FALSE	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
15	5	FALSE	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	4	FALSE	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	5	FALSE	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	5	FALSE	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	5	FALSE	29	0	0	0	1	0	0	0	0	0	0	0	2	0	1	0	0

Naive Bayes Test Results (10 Folds)
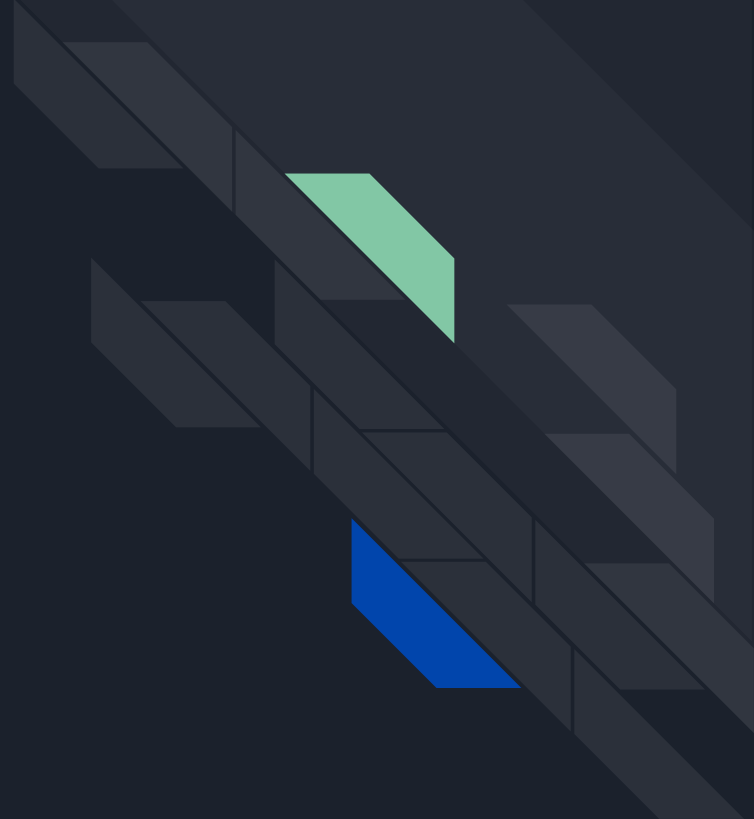
Result A

Used a list based on words that appeared 1000 times in the dataset

1 Star	2 Stars	3 Stars	4 Stars	5 Stars	
999	237	107	89	5722	**1 Star**
224	71	33	21	1293	**2 Stars**
378	118	43	49	2406	**3 Stars**
960	301	153	143	7136	**4 Stars**
13132	3876	2064	1900	105574	**5 Stars**

	1 Star	2 Stars	3 Stars	4 Stars	5 Stars
Precision	0.064	0.015	0.018	0.064	0.864
Recall	0.14	0.043	0.014	0.016	0.834

72.6591%

Overall Accuracy of Result A
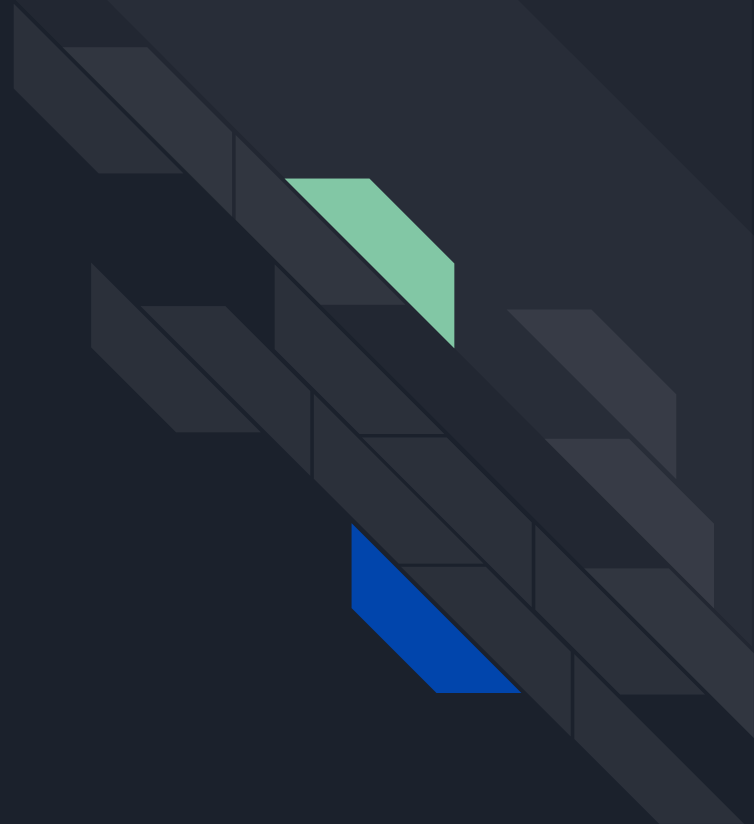
Result B

Used a custom list of 15 words
['excellent', 'however', 'cannot', 'easy', 'very', 'happy',
'really', 'not', 'love', 'good', 'poorly', 'bad', 'okay', 'great']

1 Star	2 Stars	3 Stars	4 Stars	5 Stars	
200	5	5	7	6937	1 Star
34	0	0	0	1608	2 Stars
69	0	0	5	2920	3 Stars
144	1	3	7	8538	4 Stars
1866	32	44	132	124472	5 Stars

	1 Star	2 Stars	3 Stars	4 Stars	5 Stars
Precision	0.086	0	0	0.046	0.862
Recall	0.028	0	0	0.001	0.984

84.7989%

Overall Accuracy of Result B