

Predicting Drug Review Ratings from Drug Review Text

Final Report
Lucas Reynolds

Overview

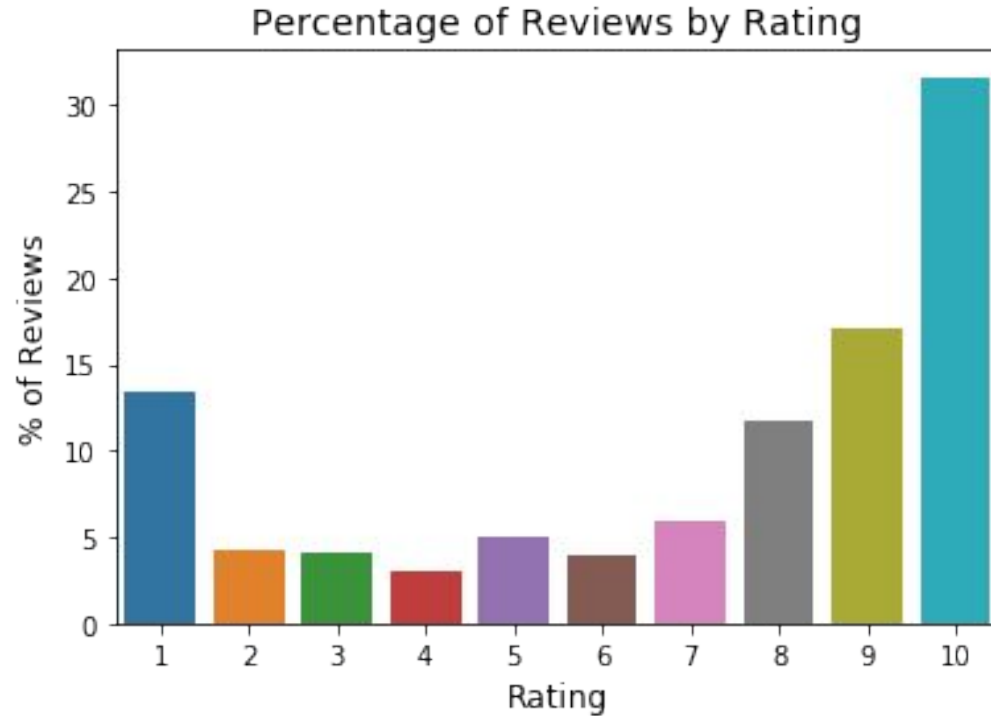
- Internet is the largest source of drug reviews
- Rating scales and requirements vary
- Building a model to predict ratings from review text would enable us to
 - Use reviews with missing ratings
 - Standardize ratings and use reviews from any source
 - Better understand factors that most influence a drug's rating

Dataset

- Over 160,000 drug reviews from Drugs.com*
- Columns:
 - Review ID number
 - Name of drug being review
 - Condition for which drug was used
 - Review text
 - Rating (from 1 to 10)
 - Review date
 - Number of times the review was voted as useful

*<https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Drugs.com%29>

Distribution of Ratings



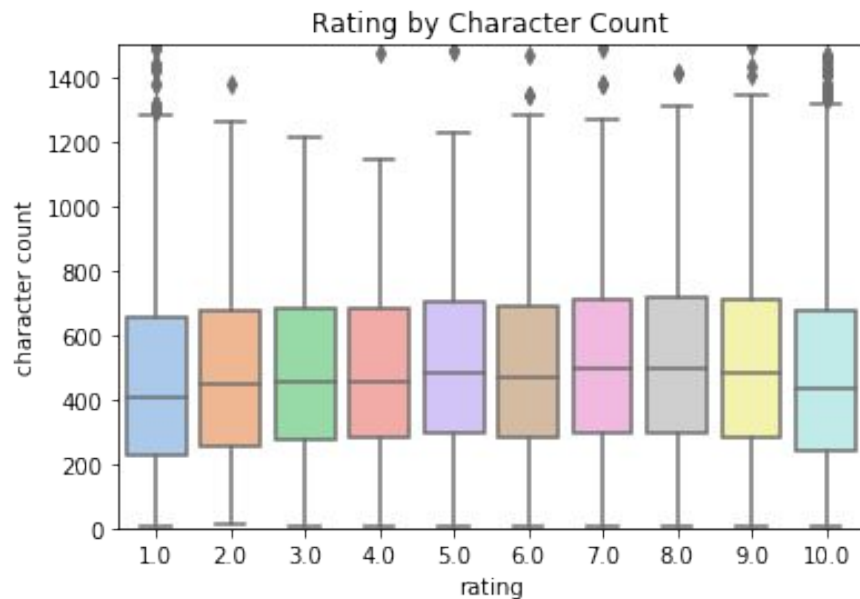
Cleaning Text

"Suboxone has completely turned my life around. I feel healthier, I'm excelling at my job and I always have money in my pocket and my savings account. I had none of those before Suboxone and spent years abusing oxycontin."



suboxon complet turn life
around feel healthier excel job
alway money pocket save
account none suboxon spent
year abus oxycontin

Review Length



Significant Correlation Coefficient: ~ 0.0267

Drug Names

5 Drugs with Largest Avg. Increase in Rating		
	Drug Name	Avg. Difference (Rating)
1	Privine	3.00
2	Zinc Oxide	3.00
3	Astelin	3.00
4	Acetaminophen / pseudoephedrine	3.00
5	Chlorpheniramine / phenylephrine	3.00

5 Drugs with Largest Avg. Decrease in Rating		
	Drug Name	Avg. Difference (Rating)
1	Systane	-4.87
2	Succinylcholine	-4.62
3	Trimethoprim	-4.25
4	Chloraseptic Sore Throat Spray	-4.00
5	Monistat 7	-3.97

Medical Condition

5 Conditions with Largest Avg. Increase in Rating		
	Medical Condition	Avg. Difference (Rating)
1	Costochondritis	3.00
2	Menopausal Disorders	3.00
3	B12 Nutritional Deficiency	3.00
4	Von Willebrand's Disease	3.00
5	Gingivitis	3.00

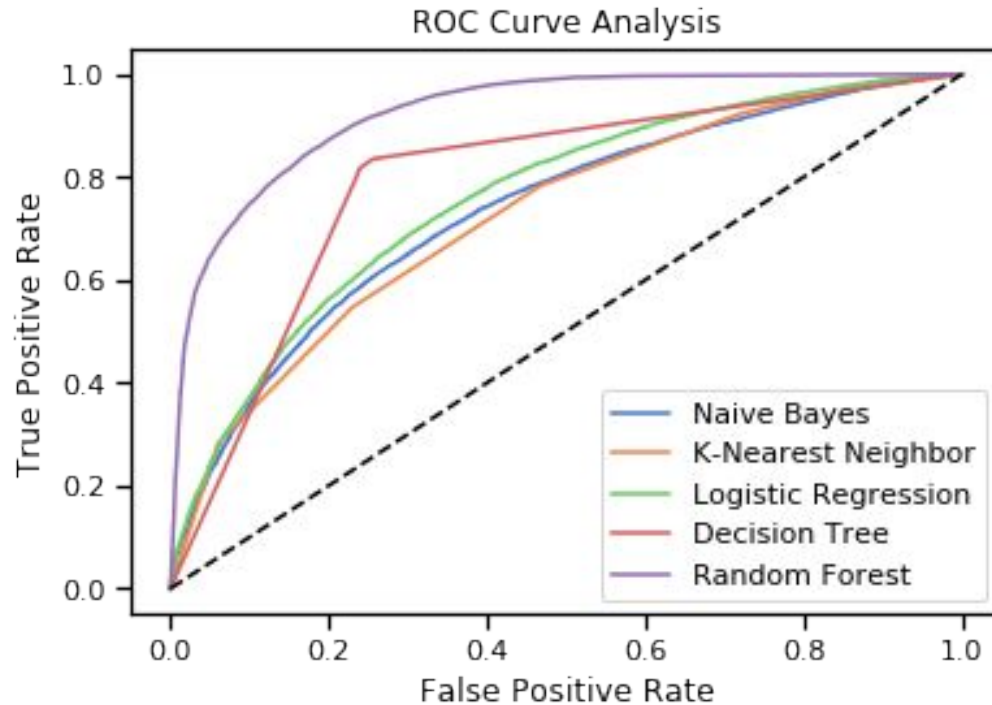
5 Conditions with Largest Avg. Decrease in Rating		
	Medical Condition	Avg. Difference (Rating)
1	Bronchospasm Prophylaxis	-3.25
2	Sore Throat	-3.11
3	Herpes Zoster, Prophylaxis	-3.00
4	Macular Edema	-2.85
5	Pancreatic Exocrine Dysfunction	-2.50

Words

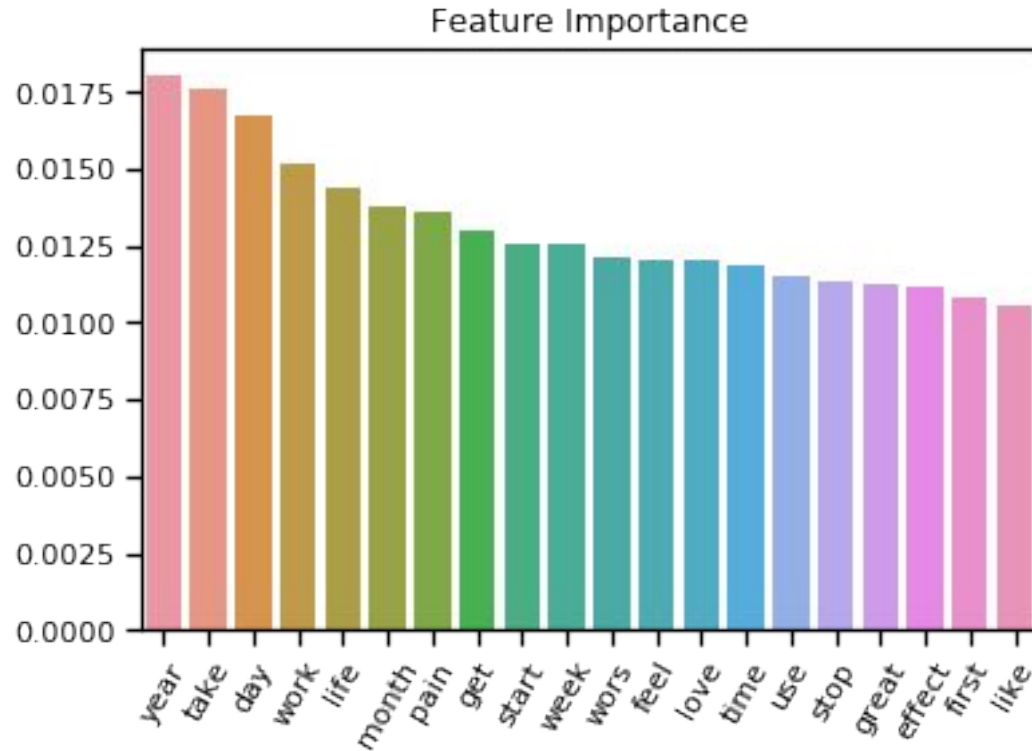
5 Words with Largest Avg. Increase in Rating		
	Words	Avg. Difference (Rating)
1	Best	1.68
2	Love	1.58
3	Great	1.15
4	Life	1.08
5	Year	0.95

5 Words with Largest Avg. Decrease in Rating		
	Words	Avg. Difference (Rating)
1	Wors	-1.93
2	Horribl	-1.75
3	Bleed	-1.52
4	Stop	-1.07
5	Switch	-1.04

Model Selection



Feature Importance



Next Steps

- Run model on drug reviews from varied sources/sites
- Try more complicated models (e.g. Recurrent Neural Network)
- Include additional features/feature engineering (e.g. conditions, drug names)