<u>Final Project NLP - UCI Medicine</u>

Bryan Sam, Caesar Pham, Marianna Carini, Karl Hickel

Code GitHub Link: https://github.com/karlhickel/NLP-project

Introduction to Dataset

The dataset of our choice for this project is a UCI medical dataset that consists of 214k rows and seven features. These features include uniqueID, drug name, condition, review, rating, date, and useful count. This dataset is a compilation of multiple types of conditions and drugs that are used to treat those conditions. The user reviews the drug and leaves a rating allowing others users to decide how helpful the review was. Overall, the dataset is very clean as is and did not require much cleaning with respect to its overall format.

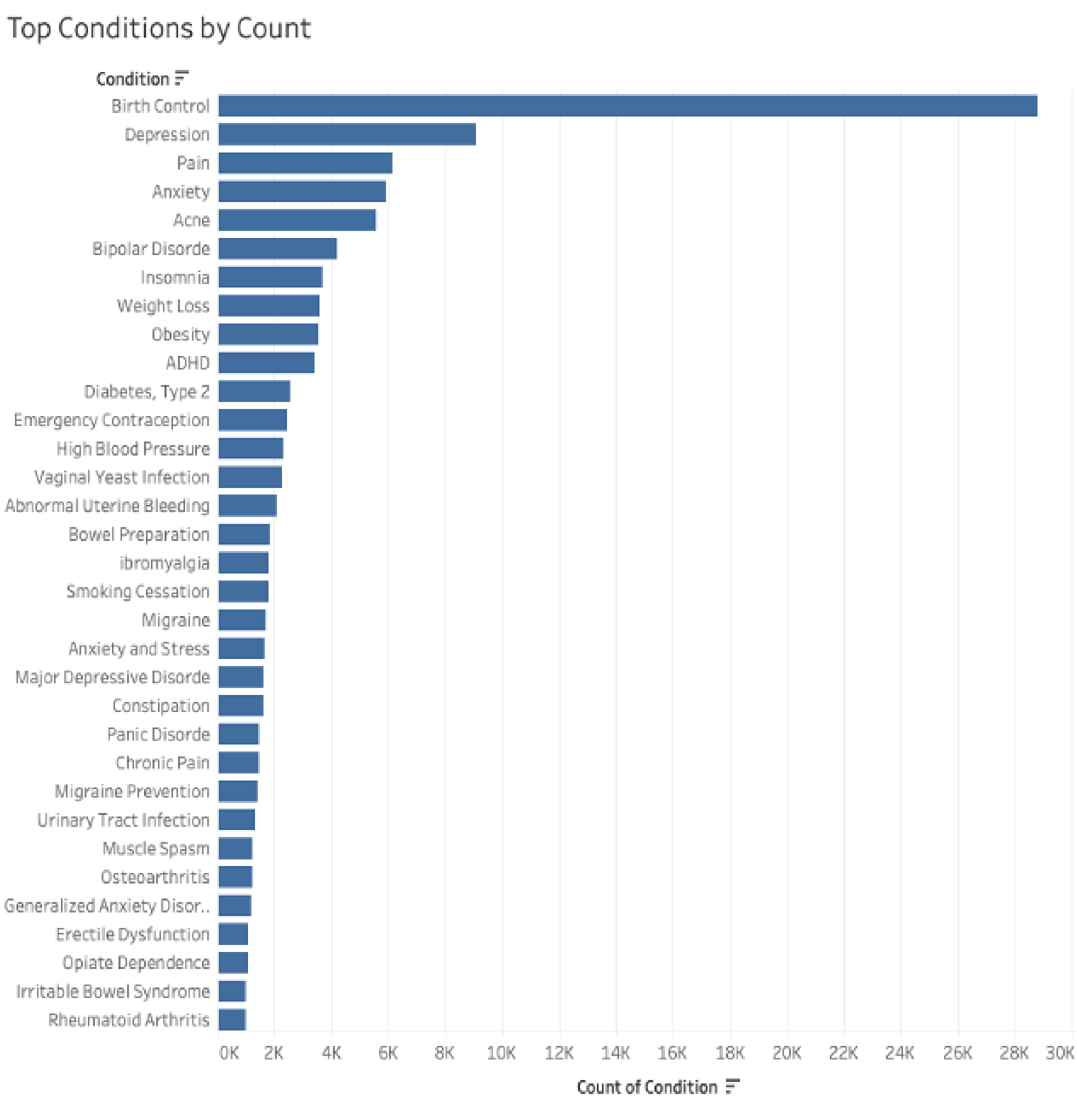
Scope of Project

The purpose of the project is to extract insights a patient would benefit from by analyzing the reviews by condition to determine which drug is effective or ineffective. In addition, models were created to classify which review is positive or negative. The scope of our project is broken into several parts and are as follows:

1. Identify most common conditions by frequency count and see average ratings over time
2. Predict whether or not the review is positive or negative
3. Establish a framework that finds the most common phrases and word(s) related to the drug (rather than the condition itself) to identify common experiences throughout reviews
    a. Ex. “Increased headaches”, “Improved Condition”

Part 1

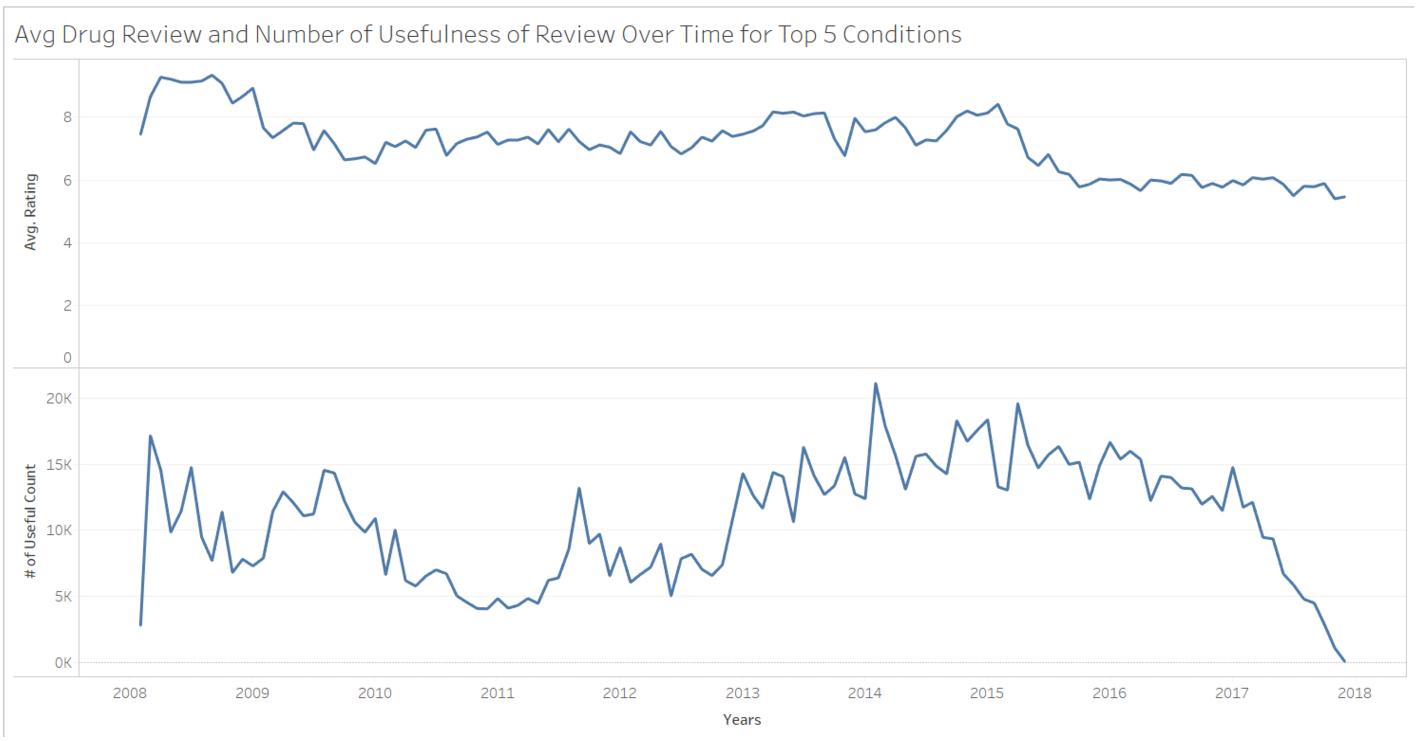
As part of our initial steps, we ensured to create some EDA that explored our dataset further identifying the most common conditions and the most highly rated drugs to treat those conditions by average user rating. Below, the Top Conditions Chart illustrates the total number of reviews by conditions.



Top Drugs for Conditions

Condition	Index	Drug Name	
Acne	1	Avar	10.000
	2	Avita	10.000
	3	Benzoyl peroxide / hyd..	10.000
	4	Benzoyl peroxide / sulf..	10.000
	5	Clindagel	10.000
	6	Ery Pads	10.000
	7	Erythromycin	10.000
	8	Fostex Medicated	10.000
	9	Hexachlorophene	10.000
	10	Norinyl 1+35	10.000
Anxiety	118	Alprazolam Intensol	10.000
	119	Diazepam Intensol	10.000
	120	Lorazepam Intensol	10.000
	121	Micrainin	10.000
	122	Pexeva	10.000
	123	Phenytoin	10.000
	124	Tranxene SD	10.000
	125	Zetran	10.000
	126	Niravam	9.800
	165	Wellbutrin	10.000
Birth Control	196	Elinest	10.000
	197	Plan B	10.000
	198	Provera	10.000
	199	Heather	9.750
	200	Ortho-Novum 7 / 7 / 7	9.667
	201	Ortho-Novum 1 / 35	8.500
	202	Femcon Fe	8.167
	203	Introvale	8.000
	204	Loestrin Fe 1.5 / 30	8.000
	205	Microgestin 1.5 / 30	8.000
Depression	368	Asendin	10.000
	369	Maprotiline	10.000
	370	Norpramin	10.000
	371	Desipramine	9.857
	372	Niacin	9.857
	373	Niravam	10.000
	374	Amitriptyline / chlordi..	9.667
	375	Serzone	9.800
	376	Amoxapine	9.500
	383	Xanax XR	10.000
Pain	473	Anaprox-DS	10.000
	474	Aspirin / caffeine	10.000
	475	Bupivacaine liposome	10.000
	476	Buprenex	10.000
	477	Children's Motrin	10.000
	478	Doans Pills Extra Stre..	10.000
	479	Exparel	10.000
	480	Feverall	10.000
	481	Lorcet	10.000
	482	Magnacet	10.000

The Top Condition chart identifies that our top conditions with the most frequency are birth control, depression, pain, anxiety and acne. The utilization of the second chart is questionable because some drugs do not account for the number of reviews rather it is only the average rating. In other words, effective drug(s) may have lower or higher average ratings because the drug(s) could have received less number of reviews compared to other drugs. The distribution of average ratings for each drug could heavily be skewed. In addition, more effective drugs, which may have lower ratings may have more reviews. This could mean the effective drugs recommended to users could be more suitable as many of similar treatments are newly introduced to the market.



Another dual time series line graph (shown above) shows users ratings of drugs and finding usefulness of reviews on drugs for treating the top 5 common conditions (birth control, depression, pain, anxiety and acne) that users face. Beginning of 2008, the average rating was moderately high until the average rating started to decrease to around the 6's from 2016 to 2018. Consequently, the number of usefulness of the reviews also decreases starting 2017 to 2018. Based on the exploratory data analysis, we subset the dataset by the top 5 conditions, which are birth control, depression, pain, anxiety, and acne, and drugs with 20 or more reviews for a narrow focus on building the predictive models and framework of finding the most common phrases and word(s) related to the drug.

Part 2

The goal of our project was to distill down a large volume of reviews into key insights for customers. Our first step in terms of natural language processing was to build a baseline model to predict whether or not a review was positive. The thought behind this was that we could use the weights of a predictive model on terms and/or ngrams to extract insights for customers. We defined a positive review to have a rating of seven or more. From the baseline, we could assess the effectiveness of subsequent models.

For our baseline model, we utilized Vader's positive and negative scores to make a rule-based classifier. We ran the entire review through Vader, and if the positive score was higher than the negative score, we predicted the review to be positive. The same logic was applied on the negative side. Our team decided precision would be the best metric to assess the effectiveness of our models. This means we would avoid telling a patient a review is positive, when in fact it is negative. Our baseline model predicted with a precision of 77%.

Next, our team improved upon that model by removing stopwords, including n-grams from one and up to 3 words to add context, and implementing regularization with both naive bayes and logistic regression. Overall, This improved our precision to 80%. Below is the model's output and the top weights.

```
Logistic Regression,Count Vecorizer Model Evaluation:

Overall Accuracy Score: {} 0.8

Confusion Matrix:
[[175  37]
 [ 49 169]]

Classification Report:
              precision    recall  f1-score   support

           0       0.78      0.83      0.80       212
           1       0.82      0.78      0.80       218

    accuracy                           0.80       430
   macro avg       0.80      0.80      0.80       430
weighted avg       0.80      0.80      0.80       430


ROC-AUC Score: 0.8003505279556863
```

Weight[7]	Feature
+0.906	love
+0.595	far
+0.410	nothing
+0.383	best
+0.350	light
… 55205 more positive …	
… 68675 more negative …	
-0.359	gained
-0.379	sex
-0.432	never
-0.451	worst
-0.538	removed

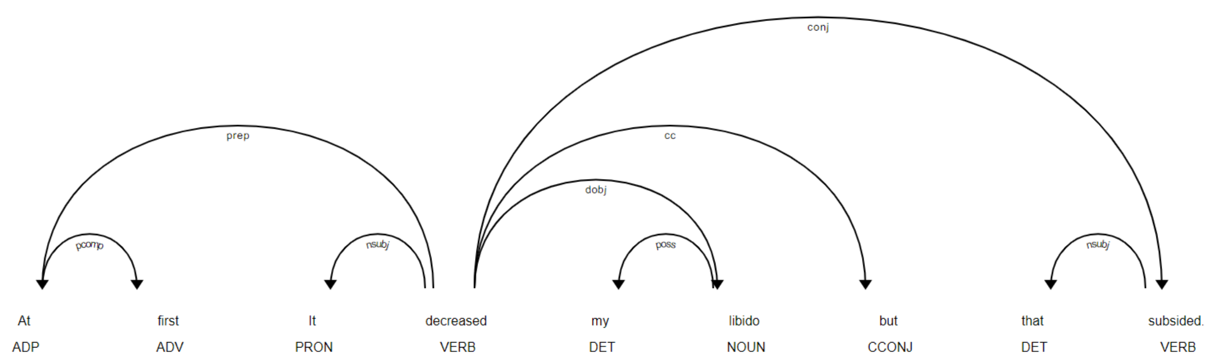
While the logistic model and the base model had relatively high precision, training the model on the entire dataset may yield an inaccurate model. Medical conditions tend to be negative in nature. Thus, traditional NLP models cannot pick up on the nuances of reviews that discuss the difficulties of a condition. For instance the review, “1 week on Zoloft for anxiety and mood swings. I take 50mg in the mornings with my breakfast. Nausea on day one but that subsided as the week went on. I get the jitters about 2 hrs after taking it followed by yawning. I feel much better though and less angry/stressed.” rates the drug as a 10/10. However, Vader rates this review as 0.141 negative and 0.068 for positivity because of the explanation of the symptoms prior to taking the drug. This leads to incorrectly classifying the review to be either positive or negative. Because of this, we looked into parsing out only the words related to the drug.

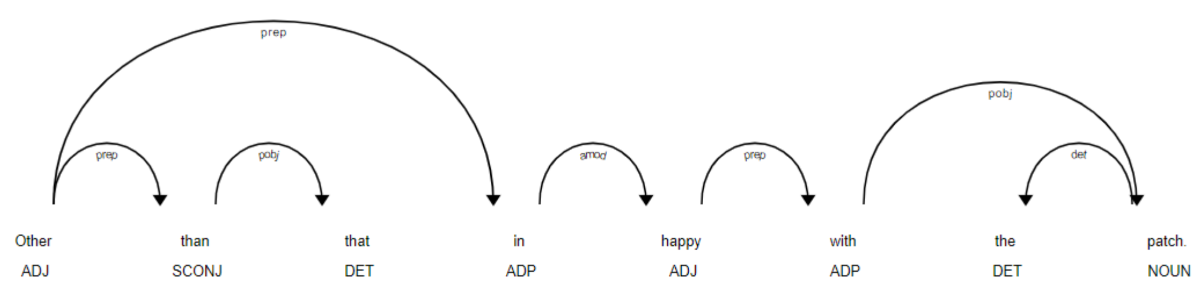
Part 3

Before establishing a framework to find the common phrases and word(s) to summarize common experiences when taking a particular drug, the reviews were cleaned and parsed for readability and to minimize words that do not create a bi-gram phrase. The text cleaning and preprocessing is the first step of our goal to create a framework. The framework of finding the most common phrases and words to summarize the common experiences when taking a drug throughout the reviews is very computationally expensive. We then focused on the top drugs based on the most common conditions to simply examine those rather than the whole dataset as that would take several hours to calculate all. spaCy is used to build the framework by understanding how the review and its sentences are structured. A review of Ortha Evra, which is used for birth control, is used as an example to examine the review and its sentences.

This is my first **ORDINAL** time using any form of birth control. I am glad I went with the patch, I have been on it for 8 months **DATE** . At first **ORDINAL** It decreased my libido but that subsided. The only downside is that it made my periods longer (5-6 days to be exact) I used to only have periods for 3-4 days **DATE** max **PERSON** also made my cramps intense for the first two days **DATE** of my period, I never had cramps before using birth control. Other than that in happy with the patch.

After rendering the Ortha Evra’s review (shown above), we noticed the end-user described her first experience using the birth control patch. She summarized how long she has been using the patch and highlighted her side effects when she using Ortha Evra. Below, we examined each rendered sentence of the Ortha Evra’s review. Here are the 2 examples:





Several associated words with the drug provide sentiment information when using the drug for treatment. Using Ortha Evra's review as an example, the user shared how the drug "decreased" her libido and how she's "happy" with Ortha Evra. The key words like "decreased" and "happy" illustrate the user's content with the drug. Rendering the reviews like the Ortha Evra's example through spaCy is used to build the framework of custom tagging of the drugs with its associated words describing the sentiment and identifying the common experiences when taking the drug. A keywords list was created to have words related to drugs and drug usages. The custom Drug Keywords list can tag specifically for drugs.

Drug Keywords List

- drug, pill, medicine, patch, injection, prescription, treatment, medication, capsule, serum, lotion, ointment, tincture, tonic, vaccination, vaccine, birth control, contraceptive, lybrel, Lybrel, Nexplanon, nexplanon, decrease, lessen, reduce, drop, diminish, dwindle, shrink, subside, cut, reduction, drop, decline, increase, improve, grow, elevate, decreased, lessened, redacted, dropped, diminished, dwindled, shrinked, subsided, cut, declined, increased, improved, grew, elevated, decreasing, lessening, reducing, dropping, diminishing, dwindling, shrinking, subsiding, cutting, declining, increasing, improving, growing, elevating

After building the Drug Keywords list, a custom drug tagging function, utilizing Drug Keywords list, was built to create a dictionary of common words based on the keyword lists. The dictionary has the key from the Drug Keywords List and its associated values. Essentially, it builds a bi-gram. Those associated values of the custom dictionary build the phrases to show sentiment and identify the common experiences that users share on particular drugs by condition. In other words, it provides context when taking a drug to treat a condition. For instance, Etonogestrel, which is used for birth control, has its own custom tagging illustrating the common experiences, side effects, and viewpoints that users share (see below).

Etonogestrel:

Key:drug, **Values:**[drive, miracle, implanon, sex, take, days, cycle, normal, chance, got, day, luck, developed, years, reacts, well, body, get, affects, delivery, effects, implanon, daily, weekly, recommend, understand, effects, would, never, went, nexplanon, exact, laughing, wrap, nausea, object, nexplanon, awful, foreign, aligns, sure, loose, wife, like, child, still, interactions, efficacy, low, outweigh, fingers, implant, good, blood, numbness, removal, month, weight, hate, nexplanon, reading, try, reviews, inserted, daughter, yrs, awful, removed, stitches, pill, months, months, dry, skin, irrititable, terrible, keep, removed, works, normalize, expected, happy, understand, associated, gain, blame, caused, convinced, included, depressed, sensitive, hormonal, recommend, got, increased, would, nov]
Key:injection, **Values:**[site, site, pain, stuff, serious, weight, started, years, depo, made, site, site, arm, every, experience, great, pleasant, months, going, try, something, day, took, spotting, spotted, taking, numbness, site, happy, get, best, depo, provera, got, tell, matchstick, changed, like, process, nothing, gain, diet, adding, healed, scarring, since, get, quickly, site, pain, inserted, arm, point, site, pain, sharp, stuff, every, aspect, basically, painless, numbing, problems, gained, control, notice]
Key:reduction, **Values:**[contraception, plus, hormones, calories, significant]
Key:decline, **Values:**[libido, complete, mood, big, get, experience, days]

Similar to Etonogestrel's custom tagging, a drug called Escitalpram is used for treating anxiety (see below).

Escitalopram:

Key:drug, **Values:**[works, miracle, sleep, things, worked, years, well, postpartum, gained, weighed, better, sure, evaluate, time, mouth, dry, chance, losing, job, effects, far, side, bed, great, great, thought, miracle, personally, made, days, miracle, anxiety, take, patient, vitamin, periods, days, posted, taking, day, tasks, could, withdrawal, increased, allowed, gain, horrible, became, helps, additional, effects, prescribed, mild, years, body, mean, something, abuse, medication, takes, results, time, get, life, stick, amazing, anxiety, dealing, life, brain, suffered, often, relationships, gave, mind, silent, life, back, made, took, make, doomed, tried, wait, saved, etc, please, australia, effect, flatulence, cheap, live, subsidised, describe, amazing, anxiety, days, brought, around, calmer, air, experience, miracle, company, lexapro, anyone, nausea, awhile, stuck, staple, anxiety, first, throughout, compared, anxiety, positive, reviews, depression, helped, take, dose, someone, small, body, subside, rating, negative, done, life, stick, lexapro, events, whole, story, sounding, miracle, system, week, getting, giving, different, feel, every, depression, heardof, working, well, others, however, realize, sleeplessness, feel, compelled, experience, given, well, others, however, realize, sleeplessness, taking, anxiety, took, chance, taking, mg, lexapro, treating, worrying, things]

Key:decrease, **Values:**[supplement, helped, attacks, days, person, still, life, without, felt, anxiety, days, feeling, weeks, libido, anxiety, medication, anxiety, social, made]

Key:reduce, **Values:**[know, ssri, acually, anxiety, suggest, notice, people, suggest, actually, tried, anxiety, effects, helped, deal, depression, still, fall]

Key:drop, **Values:**[enjoy, life, shoe, grateful, sleep]

The custom tagging helps to build context when someone is searching and taking a drug. In summary, the custom tagging of drugs by condition helps consumers to shop and purchase certain and effective drugs to treat his/her condition.

Conclusion/Future enhancements

Overall, the implementation of this framework could be useful to both consumers of the drugs and those who are prescribing them. The ability to comb through hundreds of reviews and identify key words and phrases that give some indication as to the patient's experience allows a medical professional and their patient to have a more transparent experience. Finding the right treatment can be difficult and starting a new medication regimen can be daunting; thus, our framework hopes to minimize some of the fear.

Some additional key features we would like to have added if we had more time would be further presentation enhancement features that make output more presentable. By adding a more cohesive user interface, we can have a more presentable output that is easier to comprehend. In addition to a robust user interface, it would be ideal to create a list of words that contain symptoms so we could remove words that don't make sense in our bigram pairs. Ideally we would like to grab phrases that indicate some kind effect on either the symptom of the condition or an isolated development. For instance, if the user indicates some level of fluctuation in their mood that was caused by the drug ("Increased depression", "Decreased focus"), we want that to be captured and highlighted especially if an effect is life threatening. On top of filtering important words, we believe that added context to the phrases would be beneficial as well. Our initial thoughts are that added trigrams would be of interest. Because of its computational complexity, especially with the addition of spaCy we resorted to only using bigrams. Ideally adding three words would add useful insight into the phrase. For example "drug increased headache" or "decreased my appetite" are phrases that have added to context.