Tweet Analysis to find Adverse Drug Reactions

ANKITA YADAV (MS IN BUSINESS ANALYTICS)

Table of Contents

ABSTRACT

Objective:

The abundance of text available in social media and health related forums along with the rich expression of public opinion have recently attracted the interest of the public health community to use these sources to identify ADRs from their respective drugs. Based on the intuition that patients post about Adverse Drug Reactions (ADRs) in Twitter expressing negative sentiments, we investigate the effect of sentiment analysis features in locating ADR mentions.

There are two goals of this project:

- To identify adverse drug reaction of Allegra, Zyrtec and Claritin
- Determine frequency of occurrence of common side effects

Methods:

- Twitter API was used to extract the tweets.
- The extracted tweets were then segregated based on Promotional and Non-Promotional tweets.
- Only Non- Promotional tweets were taken into consideration.
- The data cleaning and pre-processing followed by Data stemming was done in R Studio.
- Further, bag of words approach was used to filter relevant tweets from the Non-commercial tweets.
- Then the relevant tweets were classified manually as positive and negative tweets as per the effectiveness of the drug.
- From the negative tweets, tweets were identified as adverse reaction from the drug consumer
- Also, frequency of common side effects reported were determined

Result:

After successful completion of analysis of tweets, we came up with the following result:

- After extraction, *Claritin* gave 11 negative and 2 adverse reaction tweets out of the total 3439 tweets, *Zyrtec* gave 15 negative and 3 adverse reaction tweets out of the 3461 tweets, *Allegra* gave 12 negative and 4 adverse reaction tweets out of the 3461 tweets
- The most common disclosed side effect reported was drowsy/sleepy for all the three drugs

INTRODUCTION

Adverse Drug Reactions: As per FDA, "all noxious and unintended responses to a medicinal product related to any dose should be considered adverse drug reactions".

We define adverse drug reaction as "an appreciably harmful or unpleasant reaction, resulting from an intervention related to the use of medicinal product, which predicts hazard from future administration and warrants prevention or specific treatment, or alteration of the dosage regimen, or withdrawal of the product". - *US National Library of Medicine*

In Pharmacology terms, it is defined as any unexpected or dangerous reaction to a drug. An unwanted effect caused by the administration of a drug. The onset of the adverse reaction may be sudden or develop over time. – *Medicinenet.com*

The project involves identifying these Adverse Drug Reactions and the Frequency of other reactions when a consumer takes any of the following medicines: Allegra, Claritin, Zyrtec based on the Tweets by consumers over a span of last two months.

More than 50 Million Americans have experienced various types of allergies each year which makes Allergies the 6th leading cause of chronic illness in the United States. A bad drug reaction affects up to 10% of the world's population. This project is hence designed to find certain reactions that might be experienced when taking these drugs.

Twitter, one of the largest social media websites, has over 645,000,000 users (as of January 1, 2014) and grows by an estimated 135,000 users every day, generating about 9,100 tweets every second a potential gold mine of information for researchers interested in studying population trends. This is especially true given Twitter's application programming interface (API), which makes part of its data publicly available and easily accessible. A recent survey revealed that 26% of online adults discussed personal health issues and that 42% of them use social media to post or seek information about health conditions (Parker et al., 2013) [i].

With the ability to find many more adverse events than are currently found, social media like Twitter could provide an early warning detection system for surveillance of drugs, providing a safer environment for both consumers and pharmaceutical companies.

DRUG DESCRIPTION

Table A – Drug Description

Brand Name	Drug Name	First Approval by FDA
Claritin	Loratadine	November 2002
Allegra	Fexofenadine	July 1996
Zyrtec	Cetirizine	November 2007

Claritin

This medication is an antihistamine drug that treats symptoms such as itching, runny nose, watery eyes, and sneezing from "hay fever" and other allergies. It is also used to relieve itching from hives. It is also available in combination with pseudoephedrine, a decongestant, known as loratadine/pseudoephedrine.

Common side effects include sleepiness, dry mouth, and headache. Serious side effects are rare and include allergic reactions, seizures, and liver problems. Use during pregnancy appears to be safe but has not been well studied. It is not recommended in children less than two years old. It is in the second-generation antihistamine family of medication.

As a "non-sedating" antihistamine, loratadine causes less (but still significant, in some cases) sedation and psychomotor retardation than the older antihistamines because it penetrates the blood/brain barrier to a smaller extent. Other possible side effects include blurred vision, photophobia, tachycardia, difficulty in urination, hyperthermia, glaucoma, and mental confusion in the elderly

Claritin - Side Effects

Table B - Claritin Tablet Side Effects [Reference WebMD]

Common side effects (If experienced, these tend to have a Less severe expression)	• Head Pain • Headache • Drowsiness • Pink Eye • Sleepiness • Dry mouth • Cough • Dizzy • Dryness of The Nose • Earache • Excessive Sweating • Throat Dryness
Known adverse reactions (rarely experienced)	• Giant Hives • Difficulty in urination • Hyperthermia • Mental confusion • Rash • Nosebleed • Heart Throbbing or Pounding • Increased Hunger • Sun-Sensitive Skin

Tweets Count Analysis

Table C – Claritin Tweet Count Analysis:

Dates	All Tweets	Non-Promotional Tweets	Promotional Tweets
08-04-2024	95	90	5
09-04-2024	163	144	19
10-04-2024	102	88	14
11-04-2024	159	132	27
12-04-2024	143	115	28
13-04-2024	183	172	11
14-04-2024	136	104	32
15-04-2024	126	80	46
16-04-2024	110	95	15
17-04-2024	106	102	4
18-04-2024	114	109	5

Figure A – Claritin Tweet Composition
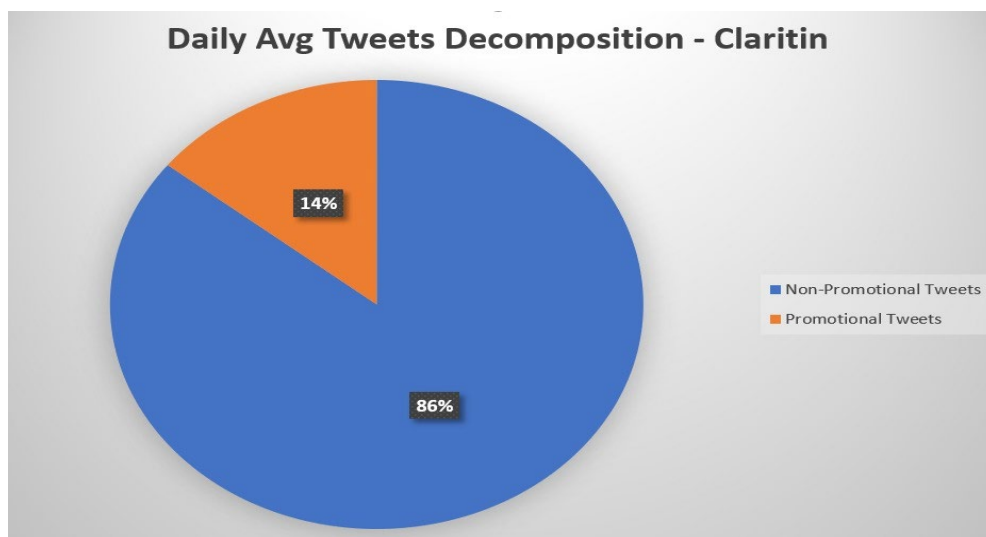
Allegra

Fexofenadine, sold under the trade name Allegra is an antihistamine used to relieve allergy symptoms such as watery eyes, runny nose, itching eyes/nose, sneezing, hives, and itching. It works by blocking a certain natural substance (histamine) that your body makes during an allergic reaction.

The most common side effect demonstrated in adults was headache, but some also experienced back and muscle pain, miosis or pinpoint pupils, nausea, drowsiness, and menstrual cramps. There have also been rare reports of anxiety and insomnia. The most common side effects demonstrated during clinical trials were cough, upper respiratory tract infection, fever, and otitis media for children ages 6 to 11 and fatigue for children ages 6 months to 5 years.

Allegra - Side Effects

Table D - Allegra Tablet Side Effects [Reference WebMD]

Common side effects (If experienced, these tend to have a Less severe expression)	- Backache - Dizzy - Drowsiness - Dry Mouth - Head Pain - Low Energy - Throat Irritation - Throwing Up
Known adverse reactions (rarely experienced)	- Acute Infection of Nose, Throat or Sinus - Hives - Hypersensitivity Drug Reaction - Life Threatening Allergic Reaction - Middle Ear Infection - Rash - Painful Periods - Indigestion - Muscle Pain

Tweets Count Analysis

Table E – Allegra Tweet Count Analysis

Dates	All Tweets	Non-Promotional Tweets	Promotional Tweets
08-04-2024	38	35	3
09-04-2024	247	199	48
10-04-2024	207	191	16
11-04-2024	284	247	37
12-04-2024	202	178	24
13-04-2024	221	199	22
14-04-2024	201	178	23
15-04-2024	233	208	25
16-04-2024	246	204	42
17-04-2024	279	246	33
18-04-2024	240	218	22

Figure B – Allegra Tweet Composition
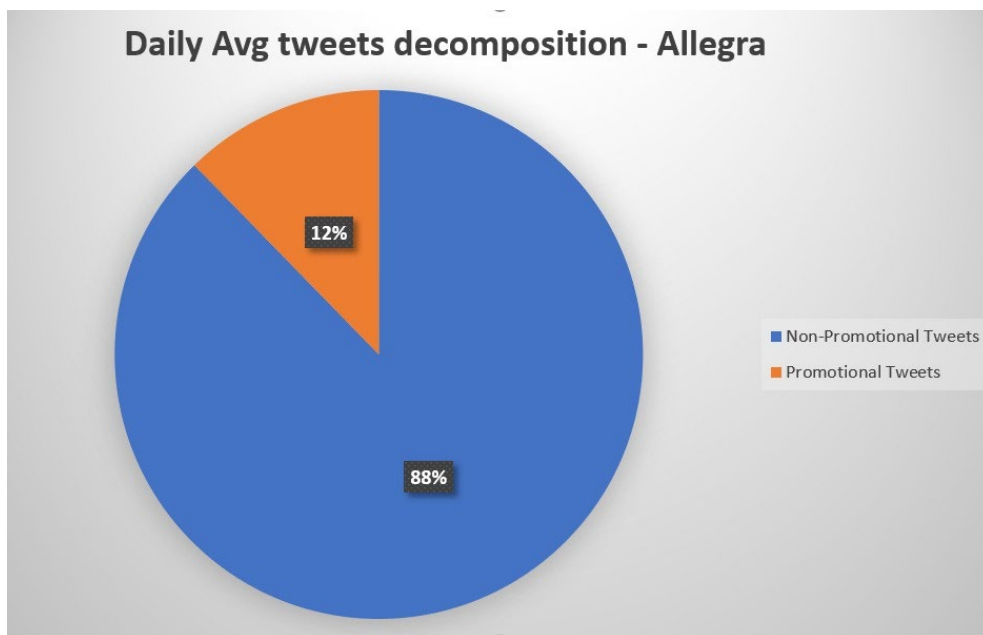
Zyrtec

Cetirizine, sold under the brand name Zyrtec among others, is a potent second-generation antihistamine used in the treatment of hay fever, allergies, angioedema, and urticaria. It acts as a selective antagonist of the histamine H1 receptor.

Cetirizine is an antihistamine used to relieve allergy symptoms such as watery eyes, runny nose, itching eyes/nose, sneezing, hives, and itching. It works by blocking a certain natural substance (histamine) that your body makes during an allergic reaction.

Commonly reported side effects of cetirizine include headache (16%), dry mouth (5.7%), drowsiness (5–20%), and fatigue (5.6%), while more serious but rare side effects include cardiac failure, tachycardia, and edema.

Zyrtec - Side Effects

Table F - Zyrtec Tablet Side Effects [Reference WebMD]

Common side effects (If experienced, these tend to have a Less severe expression)	- Dizzy - Drowsiness - Dry Mouth - Dryness of The Nose - Throat Dryness
Known adverse reactions (rarely experienced)	- Edema - tachycardia - Abnormal Heart Rhythm - Acute Pustular Eruptions on Skin - Blockage of Normal Bile Flow - Bronchospasm - Hepatitis - Life Threatening Allergic Reaction - Liver Problems - Seizures - Visible Water Retention - Chronic Trouble Sleeping - Difficult or Painful Urination - Fast Heartbeat - Increased Hunger - Involuntary Quivering - Loss of Appetite - Nervous - Nosebleed - Problems with Eyesight - Ringing in The Ears - Stomach Cramps - Sun-Sensitive Skin - Weight Gain

Tweets Count Analysis

Table G – Zyrtec Tweet Count Analysis:

Dates	All Tweets	Non-Promotional Tweets	Promotional Tweets
08-04-2024	82	75	7
09-04-2024	95	88	7
10-04-2024	97	88	9
11-04-2024	111	105	6
12-04-2024	124	121	3
13-04-2024	174	167	7
14-04-2024	107	103	4
15-04-2024	99	88	11
16-04-2024	120	89	31
17-04-2024	107	92	15
18-04-2024	79	66	13

Figure C – Zyrtec Tweet Composition
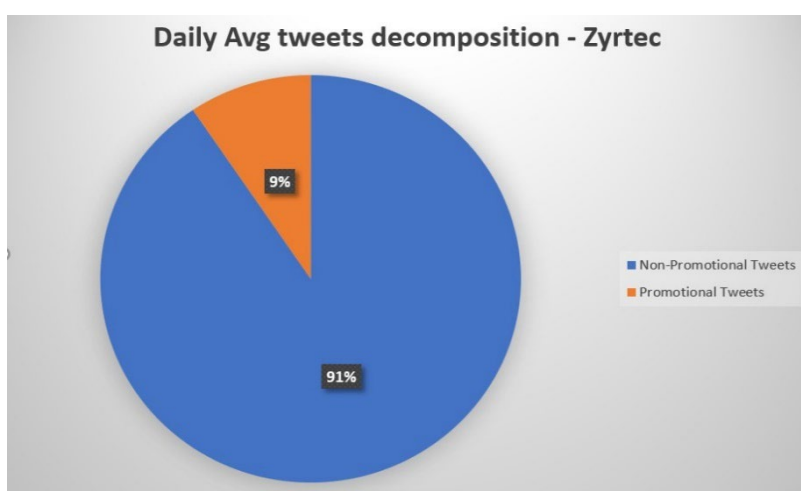
PROCESS FLOW

Figure D – Process Flow
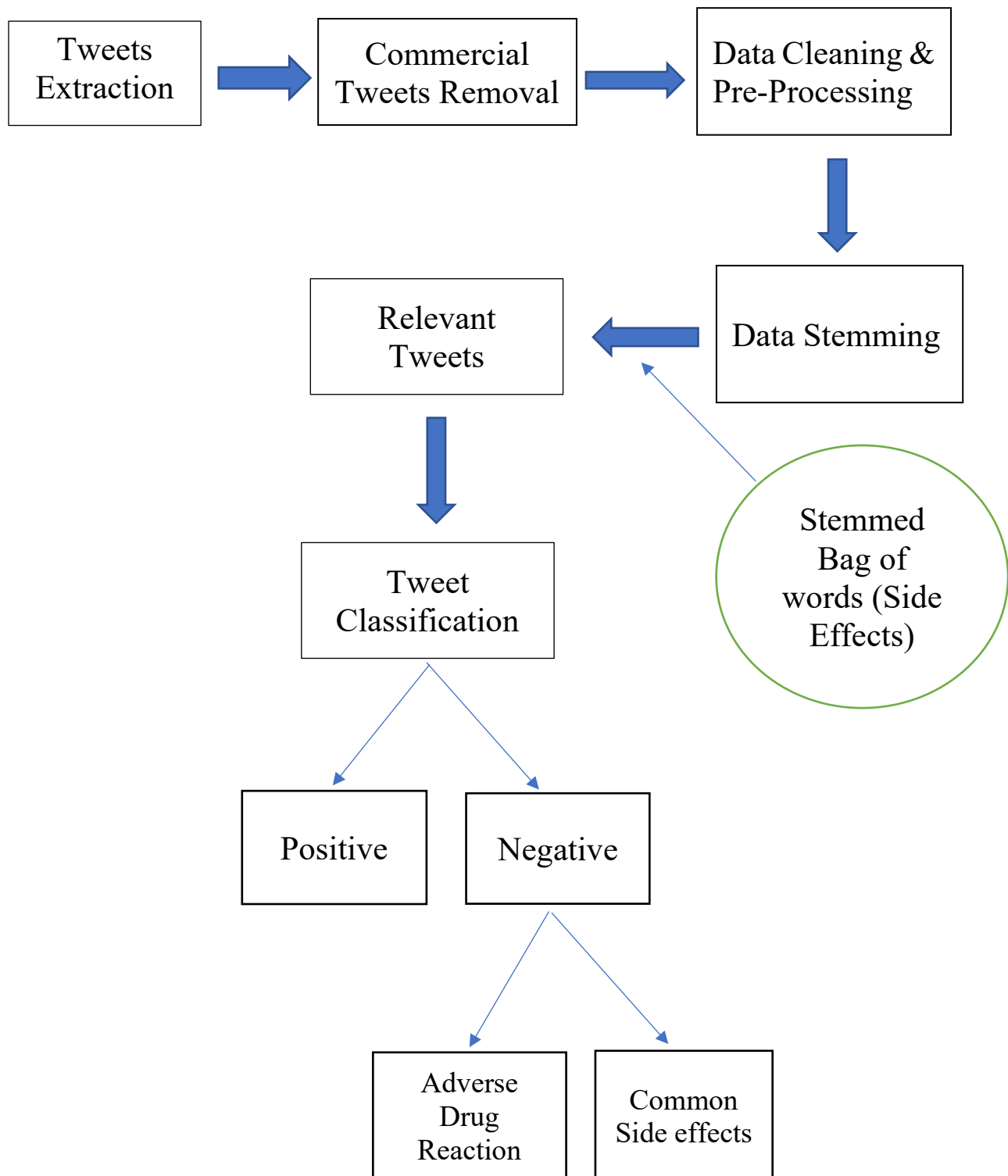
Extracting Tweets from Twitter:

We have used rtweet package to stream Twitter Data. To access the Twitter API we have used twitterR, tidyverse, httk, httr and httpuv packages. This Twitter API is used to access all the Restful API Methods of Twitter. Since this method accepts several parameters, we have used drug names in the keyword parameter to extract the tweets, n parameter to input the number of tweets to be extracted in each iteration. Since this Twitter API does not allow users to extract tweets more than two weeks, in order to collect more tweets, we have used these packages in R to manually extract one month of Twitter data from April 01, 2024 to April 28, 2024. As per the requirement, we have collected tweets of three common Anti-Allergy drugs Allegra, Claritin and Zyrtec to analyse its Adverse Drug Reactions. All the extracted tweets have been stored as csv MS Excel files.

Removing Promotional Tweets:

Marketing Tweets are the tweeted by the social media marketers of Pharma companies for promotional and commercial purposes. Since those tweets contain ad contents, coupon details, tweets creating awareness for the drugs which are irrelevant and have no value for our project, we have removed them manually by means of looking into the "source" column. This source column indicates the source from which the tweet has been posted. Also, we created bag-of-words of promotional words with which we will match and remove more promotional tweets if any. After removing these promotional and marketing tweets, we are left with relevant tweets which are tweeted by individual users who use these drugs. All these filtered tweets contain the positive and negative opinions about the drugs tweeted by the legitimate users.

Data Pre-processing:

Twitter is one of the largest social media platform which has multilingual and multi-cultural users with different backgrounds. We have considered only the English Tweets in our scope of the project. The data in the tweet generally contain many abbreviations, emoticons, emoji, typos, hashtags, URLs, mentions, grammatical errors, punctuational errors and stop words. So data pre-processing is a very crucial task before proceeding to data stemming and classification.

We have used Regular Expression to process the tweets. Regular Expression is a sequence of characters that define a search pattern. It can be either a Find operation, Find and replace operation or Find and split operation. So they are used to remove URLs, duplicate tweets, mentions, usernames, hashtag, extra newlines and all punctuations.

After applying Regular Expression, Tokenization is carried out to separate the entire text into individual words as unigram tokens which does not have any space in-between, without any special characters and converting them to lower case. Next, we have removed the "stop words" which usually refers to the most common words used in English. Moreover, we have removed duplicate tweets using duplicated method over unique id of the tweets data-frame. We did this because some of the users retweet the same tweets twice or thrice.

Data Stemming:

We did data stemming in order to map the words in the tweets with the bag of words that contains the side effects. Stemming helps to extract the root words so that those words will be the key for matching with the words in the bag of side effects. We used tm, topicmodels, tmap, qdapRegex, SnowballC, plyr and stringr packages for using Corpus function to stem the data. For example, if we stem words like "playing", "played" it will get converted to the root word "play".

We collected all the side effects of all the three drugs from WebMD and created a bag of words. And we repeated stemming process for the bag of words as well and came up with the updated stemmed bag of words for side effects.

Mapping Filtered Tweets with Bag of words (Side effects):

Before mapping we created document-term matrix of the stemmed words. Then we created a function that gets two parameters as inputs: one is the text of the filtered tweets and the other is the list of words in the bag that was scanned from the csv file created in the previous step. Then through the code it matched the tweets with any word in the bag of words and then assign a score for the number of matched bag of words with the tweet text. Once all the tweets are assigned a score based on their match, we retrieve the tweets which have a score of more than 0, i.e., the relevant tweets which have at least one side effect mentioned. We stored these tweets in a csv file.

Training of Relevant Tweets into Positive and Negative:

For the input to SVM classification of relevant tweets for sentimental analysis, we manually trained our relevant dataset of each drugs into Positive and Negative tweets. Most of the tweets were classified as positive tweets for e.g. from 152 relevant tweets of Claritin drug, 11 were classified as negative tweets.

Classification of Tweets:

From the relevant tweets csv file, we got from the previous step, we are automating a process which will classify the tweets into two categories: The one having negative opinion of users, tweets that express negative feelings and the one having positive opinion of users, tweets that express positive feelings of the drugs. For example, if a user tweets "I'm starting to feel better already after taking Allegra. I just gotta take it consistently" it is expressing a positive feeling, so its classified as a positive tweet. On the other hand, a tweet like "I took Allegra for a while but the back pain was an awful side effect" is clearly expressing a negative feeling, hence it is classified as a negative tweet.

For sentimental analysis of the relevant tweets, we used linear SVM classification. We want to be able to determine the sentiment of a tweet without any other information but the tweet text itself, hence the 'text' column is our focus. We will focus on identifying negative tweets, and hence treat neutral and positive as one class. It turns out that our dataset is unbalanced with significantly more negative than positive tweets. To explore the data we apply some crude pre-processing. We will tokenize and lemmatize using Python NLTK, and transform to

lower case. As words mostly matter in context we'll look at bi-grams instead of just individual tokens. We will build a simple, linear Support-Vector-Machine (SVM) classifier. The classifier will take into account each unique word present in the sentence, as well as all consecutive words. To make this representation useful for our SVM classifier we transform each sentence into a vector. The vector is of the same length as our vocabulary, i.e. the list of all words observed in our training data, with each word representing an entry in the vector. If a particular word is present, that entry in the vector is 1, otherwise 0.

We're now ready to fit the classifier to our data. Note here that we use the OneVsRestClassifier. This allows us to get the probability distribution over two classes. Each of these classifiers determines the probability that the data point belongs to it's corresponding class, or any of the other classes. Hence the name OneVsRest.

Screenshot 1 - Sentiment Analysis: SVM Classification

```
In [47]: sentences = count_vectorizer.transform(["zyrtec and claritin say they're non-drowsy but they work better than lunesta for sleep f
         "@rattusdingus I tried claritin but it messes with my blood pressure. Zyrtec has worked really well for me though. It only has 12
         "@killianczuba Is it that good? I feel like even non-drowsy formulas of Claritin make me sleepy, so idk. It might just be how my
         "RT @offmyplanepunk: @squidbilly929 At this point Iâ€™m just snorting Claritin and Flonase every morning and night.",
         "Need to get back on that Claritin clear allergy season is here ðŸ˜~,",
         "The weather is changing and I love it, but I don't love the allergies. Stepped outside and my eyes were watering so bad it looke
         "if our loooove is allergy then why are you my claritin",
         "@tricksythieves Claritin typically works for me if you're looking for cat allergy relief!",
         "@elevenelevenths i mean so i start to get small spots all over my body and it only occurs when iâ€™m hot so itâ€™s probably just
         ])
         clf.predict_proba(sentences)

Out[47]: array([[0.55675338, 0.44324662],
                [0.56790739, 0.43209261],
                [0.87643316, 0.12356684],
                [0.14770257, 0.85229743],
                [0.14437137, 0.85562863],
                [0.14444467, 0.85555533],
                [0.14438395, 0.85561605],
                [0.14443372, 0.85556628],
                [0.80116276, 0.19883724]])
```

- First column represents the probability of the tweet text being negative
- Second column represents the probability of the tweet text being positive
- The accuracy of the model for Claritin drug was 73%
- Mostly all the negative tweets were classified correctly

Determining Adverse Drug Reactions and Frequency of common side effects in tweets:

Out of all the negative tweets classified from the previous step, we have further classified the negative tweets into two categories: The tweets which have adverse side effects mentioned by the users and all other tweets which have common side effects mentioned by the users. As a result of this final classification, we have arrived to our goal/results – the Adverse Drug Reactions of the three drugs – Allegra, Claritin and Zyrtec.

We have determined the frequency of the common side effects tweeted by the users by means of creating document to term matrix and found the term frequency of the words that get repeated in the entire document. Then we sorted the term frequency in the descending order to get the most frequent side effects tweeted by the users.

Disclosed side-effects tweet frequency:

To get the number of tweets mentioning the common side effect for the respective drug we have automated this process using python programming. After classification of relevant tweets into positive and negative tweets, the code will match the known side effect bag of words with the negative tweets and produce list of side effects mentioned in the retrieved tweets along with the number of tweets it was mentioned in.

Screenshot 2 - Claritin: Word frequency of side effects

```
 Rows filtered by :  drowsy
['2', 'claritin a liar and the truth aint in her. the label says non drowsy so why im in this meeting sleep']
['3', 'see, i already knew my sinuses/allergies were going to try me today. so when i woke up, i took a claritin. i m drowsy, but my throa
['4', 'warm weather in the lv means my allergies are out of control. i prepared myself with a non drowsy claritin and that has seemed to b
['8', 'zyrtec and claritin say they re non drowsy but they work better than lunesta for sleep for me']
['11', ' is it that good? i feel like even non drowsy formulas of claritin make me sleepy, so idk. it might just be how my body/brain deal

 Rows numbers filtered by :  drowsy
['2', '3', '4', '8', '11']

 Rows filtered by :  sleep
['2', 'claritin a liar and the truth aint in her. the label says non drowsy so why im in this meeting sleep']
['8', 'zyrtec and claritin say they re non drowsy but they work better than lunesta for sleep for me']
['11', ' is it that good? i feel like even non drowsy formulas of claritin make me sleepy, so idk. it might just be how my body/brain deal

 Rows numbers filtered by :  sleep
['2', '8', '11']

 Rows filtered by :  headache

 Rows numbers filtered by :  headache
[]

 Rows filtered by :  blood
['9', ' i tried claritin but it messes with my blood pressure. zyrtec has worked really well for me though. it only has 12 hour versions t
>>>
```

PROCESS OUTCOME

Table H – Samples of Tweet Categories

Tweets Categories	Sample
Promotional Tweet	Printable #Coupon | Save $4.00 on Non-Drowsy Claritin® Liqui-Gels® (30 count or larger) - https://t.co/zCGtfQ80IX https://t.co/2hvxrHiiYU
Relevant Tweet	I wish I could take Benadryl so my allergies don't ruin my life today, but it makes me all sleepy and dopey, so can't take it while working. And Claritin and Flonase don't do shit.
Positive Tweet	@USNCS Ouch! I know your pain well. Ice packs work well for me. I also take allegra d the decongestant in it helps relieve the pressure. Feel better
Negative Tweet	@dizzheart I took Allegra for awhile but the back pain was an awful side effect.
Stemmed Tweet	Took allegra while back pain awful side effect

Bag of Words Sample:

Table I – Sample of Bag of Words

Marketing bag of words	Side-effect bag of words
Count	Confusion
Sweepstakes	drowsiness
vs	white
treating	infection
saved	Kidney
coupon	damage
winning	Liver
ready	Ventricular
jimmy	tachycardia
apple	Toxic
apple	epidermal
medicine	necrolysis
target	ulcers
win	Muscle
save	fatigue
prepaid	vomiting
savings	nausea
melatonin	diarrhoea
surgical	xerostomia
FlutiCare	blurred
happy	vision
prize	photophobia
forecast	urin
Marketing	hyperthermia
Johnson	glaucoma
bourbon	mental
rewarding	confusion
DayQuil	dehydrated
playlist	Hypersensitivity
itunes	abnormal
consumers	Therapeutic
google	nausea
relief	dyspnoea
	vomit
	myocardial
	infarction
	pyrexia
	convulsion
	malaise
	asthenia
	cardiac
	tachycardia
	edema
	miosis

Promotional and Non-Promotional – Classification by source

Table J - Classification by Source

Promotional Source	
	Kitty Kat Koutique
	dlvr.it
	WordPress.com
	Shoppershaul
	PCFFAutoposter
	Sprout Social
	Google
	CQ Share
	Cheap Bots, Done Quick!
	Debbie Does Coupons sp
	CouponsForYourFamily
	Nextscripts
	Triberr
	ForTheMommas
	Instagram
	NotÃcias PETROGUIA
	FeedBlitz
	leenk.me
	Symphony Tools
	Jackie Jones
	Ð–Ð¸Ð·Ð½ÑŒ Ð² ÑŠÑ‚Ð¸Ð»Ðµ
	VIP
	Buffer
	Hootsuite
	RoundTeam
	IFTTT
	Couponaholic.net
	Nursing-Resource
	Mobile Web (M2)
	Zapier.com
	Smellgists
	Medkit
	MedicalNewsToday
	SocialPilot.co
	TheCouponingCouple
	Smart Shopper Feed
	CoSchedule
	hunt4freebiescoupons
	application
	Domestic Divas Coupons
	Burlesque Name
	Twitter for Windows
	Everyone Loves Couponing
	Bitly
	Twitter Ads Composer

Non-Promotional Source	Twitter for iPhone
	Twitter for Android
	TweetCaster for Android
	Facebook
	Twitter Web Client
	Twitter Lite
	TweetDeck
	Twitter for iPad
	Flamingo for Android
	Twidere for Android #5
	Tweetbot for iÎŸS
	TwitVim
	TweetCaster for iOS
	Tweetlogix

Data Extracted from April 01, 2024 to April 28, 2024

Claritin Tweets Outcome:

Extracted Drug related tweets	3043
Non – Promotional Tweets	2702
Relevant tweets after bag of words matching	152
Negative Tweets	11
Adverse Tweets	2

Allegra Tweets Outcome:

Extracted Drug related tweets	2425
Non – Promotional Tweets	2204
Relevant tweets after bag of words matching	156
Negative Tweets	19
Adverse Tweets	4

Zyrtec Tweets Outcome:

Extracted Drug related tweets	3462
Non – Promotional Tweets	3025
Relevant tweets after bag of words matching	100
Negative Tweets	18
Adverse Tweets	3

RESULTS

Claritin Final Result

Original Tweets	Stemmed Tweets	Highlighted Side-Effect
@thecapsfan79 @capitalweather I haven't heard of that. Will have to try it. Zyrtec and Claritin always made me feel "mean" or edgy, so I switched to Allegra.	havent heard Will tri Zyrtec Claritin alway made feel mean edgi switch Allegra	edgy
Claritin a liar and the truth aint in her. The label says non drowsy so why im in this meeting sleep	Claritin liar truth aint label say non drowsi im meet sleep	drowsy
See, I already knew my sinuses/allergies were going to try me today. So when I woke up, I took a claritin. I'm drowsy, but my throat doesn't feel like it's scratchy, so mission accomplished.	See alreadi knew sinusesallergi go tri today So woke took claritin Im drowsi throat doesnt feel like scratchi mission accomplish	Drowsy, sinuses
Warm weather in the LV means my allergies are out of control. I prepared myself with a NON drowsy Claritin and that has seemed to backfire	Warm weather LV mean allergi control prepar NON drowsi Claritin seem backfir	drowsy
@ShafiqIslam0 Ugh, yeah â€" Claritin messes with my head, too. I'm busting out the flonase fo' sho.	Ugh yeah Claritin mess head Im bust flonas fo sho	headache
@elevenelevenths i mean so i start to get small spots all over my body and it only occurs when iâ€™m hot so itâ€™s probably just heat rash cause it only happens when i get too hot, it happens in the shower mostly, but also when i take allergy medicine lik	mean start get small spot bodi occur m hot s probabl just heat rash caus happen get hot happen shower most also take allergi medicin lik	skin spots
zyrtec and claritin say they're non-drowsy but they work better than lunesta for sleep for me	zyrtec claritin say theyr non drowsi work better lunesta sleep	drowsy
@rattusdingus I tried claritin but it messes with my blood pressure. Zyrtec has worked really well for me though. It only has 12 hour versions tho unlike claritioin so gotta pop it every morning and evening ;9	tri claritin mess blood pressur Zyrtec work realli well though hour version tho unlik clarit gotta pop everi morn even	blood pressure
Dear #Bayer When did your ðŸ'"ðŸ''ðŸ'œðŸ'œðŸ‡®ðŸ‡¹ðŸ'¦controlled pharmacists begin adding a derivative of Benadryl to Claritin?Theyâ€™re aware Iâ€™m allergic and recent batch gave me full body skin rash until I stopped taking it. ðŸ—•ðŸ'» #Facts https://t.co/JB4kOXZ0KE	Dear Bayer When control pharmacist begin ad deriv Benadryl ClaritinThey re awar m allerg recent batch gave full bodi skin rash stop take Fact https tcoJBkOXZKE	skin rash
@killianczuba Is it that good? I feel like even non-drowsy formulas of Claritin make me sleepy, so idk. It might just be how my body/brain deal with any bit of unhealth.	good feel like even non drowsi formula Claritin make sleepi idk might just bodybrain deal bit unhealth	drowsy

Allegra Final Result

Original Tweets	Stemmed Tweets	Highlighted Side-Effect
Switched from Claritin to Allegra...world of difference, though now I can't sleep. ðŸ˜	switch claritin allegraworld differ though now cant sleep	sleep
@rbchoopmas @simplealice241 if the testimony doesn't put me to sleep, Allegra certainly will. Lol.	testimoni doesnt put sleep allegra certain will lol	sleep
@JaySwann_ I donâ€™t even think Allegra comes in a drowsy type. So no.	don t even think allegra come drowsi type	drowsy
While Allegra D gives me life at the onset of spring, itâ€™s got to be knocking years off with the increased body temperature, dry mouth, and ruthless headaches. #fineprint #spring	allegra d give life onset spring s got knock year increas bodi temperatur dri mouth ruthless headach fineprint spring	increased body temperature, dry mouth, ruthless headache
@dizzheart I took Allegra for a while but the back pain was an awful side effect.	took allegra awhil back pain aw side effect	back pain

Zyrtec Final Result

Original Tweets	Stemmed Tweets	Highlighted Side-Effect
I took a 24 hour Zyrtec 12 hours ago. Why do I feel like my eyes are gonna itch outta my head	took hour zyrtec hour ago feel like eye gonna itch outta head	itchy eyes
I haven't heard of that. Will have to try it. Zyrtec and Claritin always made me feel "mean" or edgy, so I switched to Allegra.	havent heard will tri zyrtec claritin alway made feel mean edgi switch allegra	edgy
Took a Zyrtec, congestion is better but now I feel a bit lightheaded and woozy.	took zyrtec congest better now feel bit lighthead woozi	lighthead, woozy
singswell Will do! Iâ€™m trying really hard to stay up a little but I took a Zyrtec for my allergies and Iâ€™m starting to feel a little sleepy ðŸ˜©	singswel will m tri realli hard stay littl took zyrtec allergi m start feel littl sleepi	sleepy
capitalweather Zyrtec puts me to sleep but itâ€™s great for allergies. Allegra is best for me.	capitalweath zyrtec put sleep s great allergi allegra best	sleepy
When you got severe pollen allergies but recently learned that antihistamines can make yo coochie dry so you round here sneezing and scratching yo ears and throat..scared to take a Zyrtec ðŸ˜˜,ðŸ˜˜,ðŸ˜˜,	got sever pollen allergi recent learn antihistamin can make yo coochi dri round sneez scratch yo ear throatscar take Zyrtec	dry vagina
let me know if you find anything that works well; for me claritin is ok for pet allergies, and zyrtec has worked decently for me but it makes me drowsy, along with benadryl and sinus headaches make me want to smash my head like a melon	let know find anyth work well claritin ok pet allergi zyrtec work decent make drowsi along benadryl sinus headach make want smash head like melon	drowsy
I donâ€™t care what Zyrtec says. That shit is not non drowsy. I feel so doped up right now.	don t care zyrtec say shit non drowsi feel dope right now	drowsy
feel like pure shit zyrtec lied when it said it wouldnt make me drowsy https://t.co/ErdrQoQkpK	feel like pure shit zyrtec lie said wouldnt make drowsi https tcoerdrqoqkpk	drowsy
Be aware for some of us unlucky folk, Zyrtec causes bad stomach aches.	awar us unlucki folk zyrtec caus bad stomach ach	stomach ache
RT @SoppphieWright: In order to get good sleep I have to take a Zyrtec, acid reducer, and have a neck pillow. I may be an old man stuck iâ€¦	rt order get good sleep take zyrtec acid reduc neck pillow may old man stuck	sleep
zyrtec and claritin say they're non-drowsy but they work better than lunesta for sleep for me	zyrtec claritin say theyr non drowsi work better lunesta sleep	drowsy
I finally went to my moms house and got grava leaves for my Zyrtec rash breakout. Thank god!!!!	final went mom hous got grava leav zyrtec rash breakout thank god	rash

Claritin Side-effects Tweets Frequency

Side Effect	# of Negative Tweets
Drowsy	5
Sleepy	3
Rash	2
Edgy	1
Scratchy	1
Skin spots	1
Blood pressure	1

Allegra Side-effects Tweets Frequency

Side Effect	# of Negative Tweets
Sleepy	2
Drowsy	1
Headache	1
Increase body temperature	1
Back pain	1
Dry mouth	1

Zyrtec Side-effects Tweets Frequency

Side Effect	# of Negative Tweets
Sleepy	7
Drowsy	4
Itchy	2
Edgy	2
Lightheaded	1
Dizzy	1
Scratchy eyes	1
Stomach ache	1
Rash	1

Adverse Drug Reaction Extraction:

The ADRs extraction is based on the initial drug-related tweets. For each of the 3 drugs a bag of Adverse Drug Reaction was created from WebMD. This bag of words was then matched with the processed tweets and the matching tweets were separated from the original dataset. These tweets were then manually classified into positive and negative based on the sentiment of the tweets related to the drug. Adverse drug reactions were then identified from the negative tweets.

Below is the list of ADRs of each drug separated into Disclosed and Undisclosed ADRs:

Disclosed ADRs:

Drugs	Adverse Drug Reactions
Claritin	• Spots and Rashes
Allegra	N/A
Zyrtec	N/A

Undisclosed ADRs:

Drugs	Adverse Drug Reactions
Claritin	• Blood Pressure Issue
Allegra	• Back Pain
Zyrtec	• Stomach Ache • Rashes • Dry Vagina

Most Frequent Drug Reactions:

The negative tweets were further processed to remove stop words. A document-term matrix was then generated to calculate frequency of each term. Drug reaction related words were then separated and sorted in descending order to figure out the most frequent reactions.

Drugs	Most Frequent Reactions
Claritin	• Drowsy/ Sleepy
Allegra	• Drowsy/Sleepy
Zyrtec	• Drowsy/Sleepy

FUTURE SCOPE:

Insights – Twitter can be recognized as one of the reliable and legitimate sources to find Adverse Drug Reactions.

Extract more tweets from Twitter for processing a large amount of data which can be used to train the classifier more and improve the accuracy of classification.

APPENDIX:

CODE: Tweets Extraction

Screenshot 3 – Code for Tweet Extraction:

```r
install.packages("twitteR") #installs TwitteR library (twitteR) #loads TwitteR
library(twitteR)
library(rtweet)
library(tidyverse)
install.packages('tidyverse')
install.packages("httk")
install.packages("httr")
install.packages("httpuv")
install.packages("rtweet")

consumer_key = 'QjA2bFOSHnFeQ4B4pNKWOkBRA'
consumer_secret = 'Dx4wxueiWbn8TlErF5yYAjFrNViZ3mJZNt4H6DrFSmAg4P7o55'
access_token = '138107497-NO02BdvIl9Ym3D6teY4svo8uaQ2v0yfWiDUM5dfA'
access_token_secret = '1n5EIF0CXJOIfJO5T4KRJYTRWhe4eQv8p32FoLJAgpKNU'


appname <- "Twitter_ADBMS_TS"
token <- create_token(app =appname, consumer_key, consumer_secret,set_renv = TRUE)


lrt <- lapply(
  c("Claritin", "#Claritin"),
  search_tweets,
  n = 7000,lang = "en"
)

rt9 <- do_call_rbind(lrt)
View(rt9)
rt9 <- as.data.frame(rt9)
str(rt9)
write_as_csv(rt9,'/Users/akashadlakha/Desktop/UIC/Spring 18/Information Stratergy and Policy/Project/Claritin_new25.csv')
rt9 <- rt9[!duplicated(rt9$text),]
```

CODE: For Cleaning, stemming and document term frequency generator.

```r
data <- read.csv(file.choose())

#Stemming and Frew words
docs <- Corpus(VectorSource(data$text))
docs <-tm_map(docs,content_transformer(tolower))

toSpace <- content_transformer(function(x, pattern) { return (gsub(pattern, " ", x))})

docs <- tm_map(docs, toSpace, "-")
docs <- tm_map(docs, toSpace, "'")
docs <- tm_map(docs, toSpace, "'")
docs <- tm_map(docs, toSpace, "•")
docs <- tm_map(docs, toSpace, """)
docs <- tm_map(docs, toSpace, """)
docs <- tm_map(docs, toSpace, "#")
docs <- tm_map(docs, toSpace, "<")
docs <- tm_map(docs, toSpace, ">")
docs <- tm_map(docs, toSpace, "@\\w+ *")
docs <- tm_map(docs, toSpace, "&\\w+ *")
docs <- tm_map(docs, toSpace, "[^\x01-\x7F]")
docs <- tm_map(docs, toSpace, "(f|ht)tp(s?)://(.*)[.][a-z]+")
docs <- tm_map(docs, toSpace, ":")
```

```
data$text[1]
writeLines(as.character(docs[[1]]))

#remove punctuation
docs <- tm_map(docs, removePunctuation)
#Strip digits
docs <- tm_map(docs, removeNumbers)
#remove stopwords
docs <- tm_map(docs, removeWords, stopwords("english"))
#remove whitespace
docs <- tm_map(docs, stripWhitespace)
#Good practice to check every now and then
writeLines(as.character(docs[[1]]))
#Stem document
docs1 <- docs

docs <- tm_map(docs,stemDocument, language = "english")
writeLines(as.character(docs[[1]]))

#Create document-term matrix
dtm <- DocumentTermMatrix(docs)
#convert rownames to filenames
#rownames(dtm) <- filenames
#collapse matrix by summing over columns
freq <- colSums(as.matrix(dtm))
#length should be total number of terms
length(freq)
#create sort order (descending)
ord <- order(freq,decreasing=TRUE)
#List all terms in decreasing order of freq and write to disk
freq[ord]

write.csv(freq[ord],"Result.csv")   #CSV file with frequency of each word
                                            #in stemmed form.
```

CODE: For Matching of words with the Bag of Words

```
word.match <- function(sentences,list.words){

 scores<-laply(sentences,function(sentence,list.words){
   sentence<-gsub('[[:punct:]]',"",sentence)
   sentence<-gsub('[[:cntrl:]]',"",sentence)
   sentence<-gsub('\\d+',"",sentence)

   word.list<-str_split(sentence,'\\s+')
   words<-unlist(word.list)
   pos.matches<-match(words,list.words)
```

```
  pos.matches<-!is.na(pos.matches)
  score<-sum(pos.matches)
  return(score)
},list.words)
scores.df<-data.frame(score=scores,text=sentences)
return(scores.df)
}


data <- read.csv(file.choose())
list.words<-scan('bag_of_words.csv', what = 'character', comment.char = '')

test.score<- word.match(claritin$text,list.words)

View(test.score)

Result <- test.score[test.score$score>0,]
write.csv(Result, "Result.csv")
```

CODE: Automated process - Get mentioned known side effect with Number of tweets

Screenshot 4: Code for automating the process of extracting known side effects

```
 1  import csv
 2  class FilterDataFrame:
 3      def filter_data_frame(self, filters):
 4
 5          for fil in filters:
 6              filtered_rows = []
 7              row_numbers = []
 8              with open('Claritin_Negative_Tweets_Classified.csv', "r") as f:
 9                  reader = csv.reader(f)
10
11                  for row in reader:
12                      for col in row:
13                          if fil in col:
14                              filtered_rows.append(row)
15                              row_numbers.append(row[0])
16                              break
17
18              print("\n Rows filtered by : ", fil)
19              for filtered_row in filtered_rows:
20                  print(filtered_row)
21              print("\n Rows numbers filtered by : ", fil)
22              print(row_numbers)
23  if _name_ == "_main_":
24      filterObj = FilterDataFrame()
25      filters = ['drowsy', 'sleep', 'headache', 'blood','Anemia','Angioedema','Bone','fractures','Blood','clots',
26      'Confusion','drowsiness','white','infection','Kidney','damage','Liver',
27      'Ventricular','tachycardia','Toxic','epidermal','necrolysis','ulcers','Muscle','fatigue','vomiting','nausea','diarrhoea',
28      'xerostomia','blurred','vision','photophobia','urin','hyperthermia','glaucoma','mental','confusion','dehydrated','Hypersensitivity',
29      'abnormal','Therapeutic','nausea','dyspnoea','vomit','myocardial','infarction','pyrexia','convulsion','malaise','asthenia',
30      'cardiac','tachycardia','edema','miosis','menstrual','anxiety','insomnia','otitis','infection','coochie','bleed','blood','pressure',
31      'rash','spots']
32      filterObj.filter_data_frame(filters)
```

REFERENCES:

i. https://open.fda.gov/drug/event/

ii. https://www.fda.gov/downloads/drugs/guidances/ucm073087.pdf

iii. https://ieeexplore.ieee.org/document/7752365/

iv. https://www.sciencedirect.com/science/article/pii/S1532046416300508

v. https://www.merckmanuals.com/home/drugs/adverse-drug-reactions/overview-of-adverse-drug-reactions

vi. http://www.nactem.ac.uk/biotxtm2014/papers/Ginnetal.pdf

vii. https://www.webmd.com/

viii. https://www.figure-eight.com/discovering-drug-side-effects-with-crowdsourcing/