

Medicine Recommendation System

Introduction and Background of the Project:

There are Multiple Drugs in the Market for every health issue and sometimes it is very difficult for the patients to sit for a long time in the waiting room until they get their turn for a doctor's appointment and getting some medicines suggested by doctors. Now we can get rid of all these processes and make it easy for both doctors and patients by developing a Medicine Recommendation System using Machine Learning and recommendation techniques.

This Project focuses on Medical Recommendation System which can predict accurate medical needs or medicines required for a patient based on the user (Patient) inputs about disease or symptom and these recommendations are mainly designed to use various filtering methods to recommend a specific product to its users. Using NLP techniques, we can learn the utility of items to each user which helps in predicting the medication and is trained on unsupervised learning.

Statement of the Project Problem:

Since Medicines are undoubtedly, one of the saviors of humankind a medical recommendation system is an essential tool for patients and as well as doctors who work 24*7 under heavy pressure, and patients will also have a second opinion on the medicines which are recommended by another doctor and thus could help in improving the productivity of clinical treatment by recommending standard medicines and reducing personal cost.

Review of Literature:

In past research there are many recommendation systems in health care, one of them is recommending doctors and medicines using review mining, this recommendation system will suggest doctors and medicines to users. For this, they used an SVM (support vector machine), VSM (vector space model) based health recommendation system (HRS) is discussed for the user that contains a web-based assessment system using review mining [1]. Another recommendation system is like Medical Recommendation systems this type of Recommendation system will help to explore potential knowledge from diagnosis history records and help doctors to prescribe medication correctly and for creating this strong Recommendation system they used the Naïve Bayesian Classifier ML method [2]. Recommendation systems that I am going to create are related to healthcare and little related to two previous works that I mentioned.

Objectives of the Study:

As there are many medicines in the market and it is very difficult for the patients which one to use for any health issue. To determine this factor, I am going to design a recommendation system that will help both patients to take the right medicine for any health issue and with this system, they can get some knowledge on medicines.

Data Collection:

Dataset is collected from UCI Machine Learning Repository. My dataset consists of 107532 records with 7 columns unique ID, condition, rating, drug name, review, date, useful count. Condition column represents the health issue of patient and in rating columns says how much patient is suffering from that

health issue and then the third column is the drug name for that health issue along with rating. Review column tells how patients feel about that medicine and useful counts tells how many people feel that review helps them.

	uniqueID	drugName	condition	review	rating	date	usefulCount
0	163740	Mirtazapine	Depression	"I've tried a few antidepressants over th...	10	28-Feb-12	22
1	206473	Mesalamine	Crohn's Disease, Maintenance	"My son has Crohn's disease and has done ...	8	17-May-09	17
2	159672	Bactrim	Urinary Tract Infection	"Quick reduction of symptoms"	9	29-Sep-17	3
3	39293	Contrave	Weight Loss	"Contrave combines drugs that were used for al...	9	5-Mar-17	35
4	97768	Cyclafem 1 / 35	Birth Control	"I have been on this birth control for one cyc...	9	22-Oct-15	4
...
53761	159999	Tamoxifen	Breast Cancer, Prevention	"I have taken Tamoxifen for 5 years. Side effe...	10	13-Sep-14	43
53762	140714	Escitalopram	Anxiety	"I've been taking Lexapro (escitalopram)...	9	8-Oct-16	11
53763	130945	Levonorgestrel	Birth Control	"I'm married, 34 years old and I have no ...	8	15-Nov-10	7
53764	47656	Tapentadol	Pain	"I was prescribed Nucynta for severe neck/shou...	1	28-Nov-11	20
53765	113712	Arthrotec	Sciatica	"It works!!!"	9	13-Sep-09	46

107532 rows × 7 columns

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

Exploratory data analysis :

This below screenshot tells us the summary statistics of each columns in our dataset.

```
df.describe()
```

	uniqueID	rating	usefulCount
count	107532.000000	107532.000000	107532.000000
mean	116386.701187	6.976900	27.989752
std	67017.428260	3.285192	36.172665
min	0.000000	1.000000	0.000000
25%	58272.000000	4.000000	6.000000
50%	116248.500000	8.000000	16.000000
75%	174588.000000	10.000000	36.000000
max	232284.000000	10.000000	949.000000

There are 590 null values in condition column.

```
print(df.isnull().sum())
```

```
uniqueID      0
drugName      0
condition     590
review        0
rating        0
date          0
usefulCount   0
dtype: int64
```

Unique values in condition column.

```
df["condition"].value_counts()
```

```
Birth Control      19296
Depression         6190
Pain               4200
Anxiety           3816
Acne              3694
...
Head and Neck Cance      2
Peripheral Arterial Disease      2
Tympanostomy Tube Placement Surgery      2
Chronic Inflammatory Demyelinating Polyradiculoneuropathy      2
Neutropenia           2
Name: condition, Length: 708, dtype: int64
```

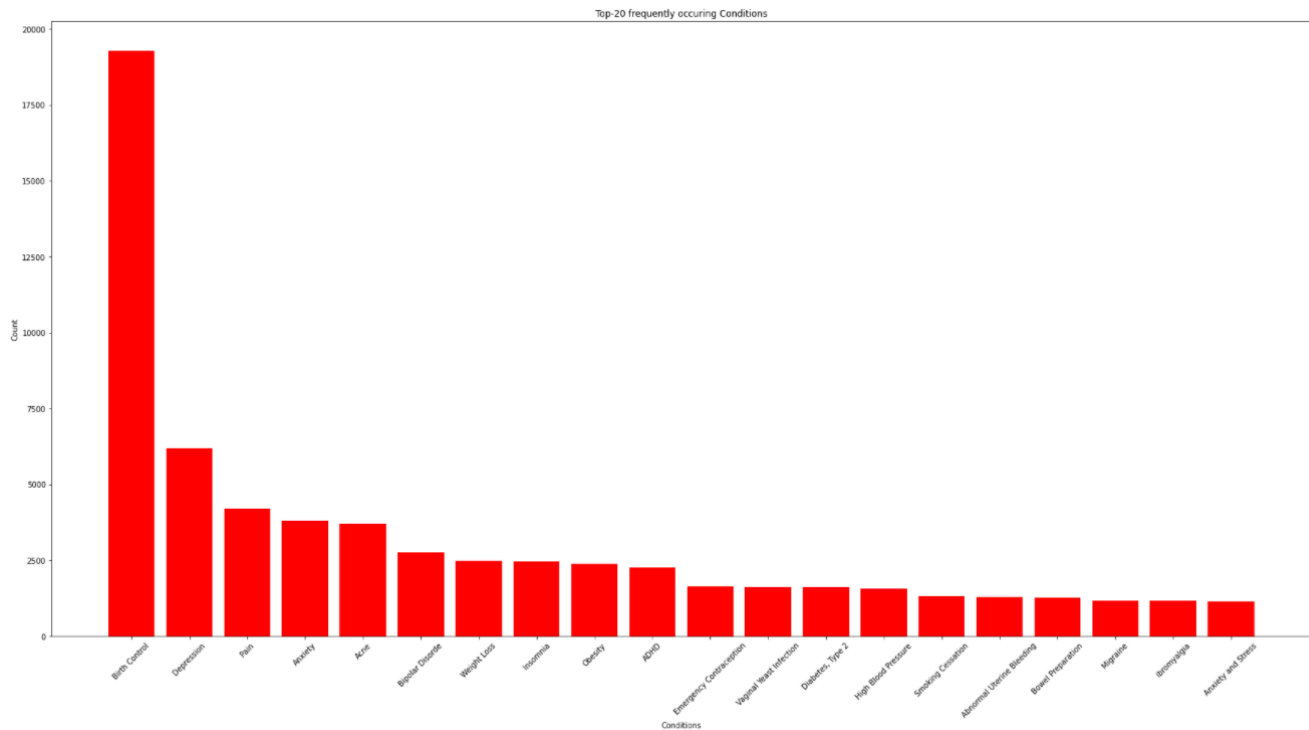
I have removed Span names rows from the condition column because there won't be any use in our analysis.

```
df=df[df["condition"].str.contains("span")==False]
df
```

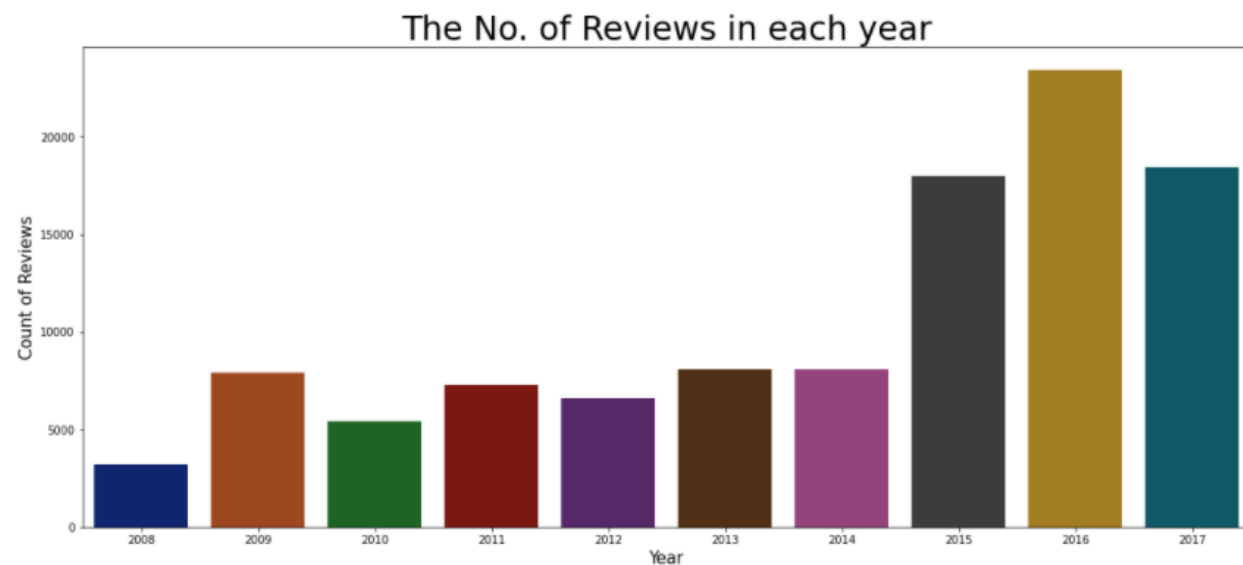
	uniqueID	drugName	condition	review	rating	date	usefulCount
0	163740	Mirtazapine	Depression	"I've tried a few antidepressants over th...	10	28-Feb-12	22
1	206473	Mesalamine	Crohn's Disease, Maintenance	"My son has Crohn's disease and has done ...	8	17-May-09	17
2	159672	Bactrim	Urinary Tract Infection	"Quick reduction of symptoms"	9	29-Sep-17	3
3	39293	Contrave	Weight Loss	"Contrave combines drugs that were used for al...	9	5-Mar-17	35
4	97768	Cyclafem 1 / 35	Birth Control	"I have been on this birth control for one cyc...	9	22-Oct-15	4
...
53761	159999	Tamoxifen	Breast Cancer, Prevention	"I have taken Tamoxifen for 5 years. Side effe...	10	13-Sep-14	43
53762	140714	Escitalopram	Anxiety	"I've been taking Lexapro (escitalopra...	9	8-Oct-16	11
53763	130945	Levonorgestrel	Birth Control	"I'm married, 34 years old and I have no ...	8	15-Nov-10	7
53764	47656	Tapentadol	Pain	"I was prescribed Nucynta for severe neck/shou...	1	28-Nov-11	20
53765	113712	Arthrotec	Sciatica	"It works!!!"	9	13-Sep-09	46

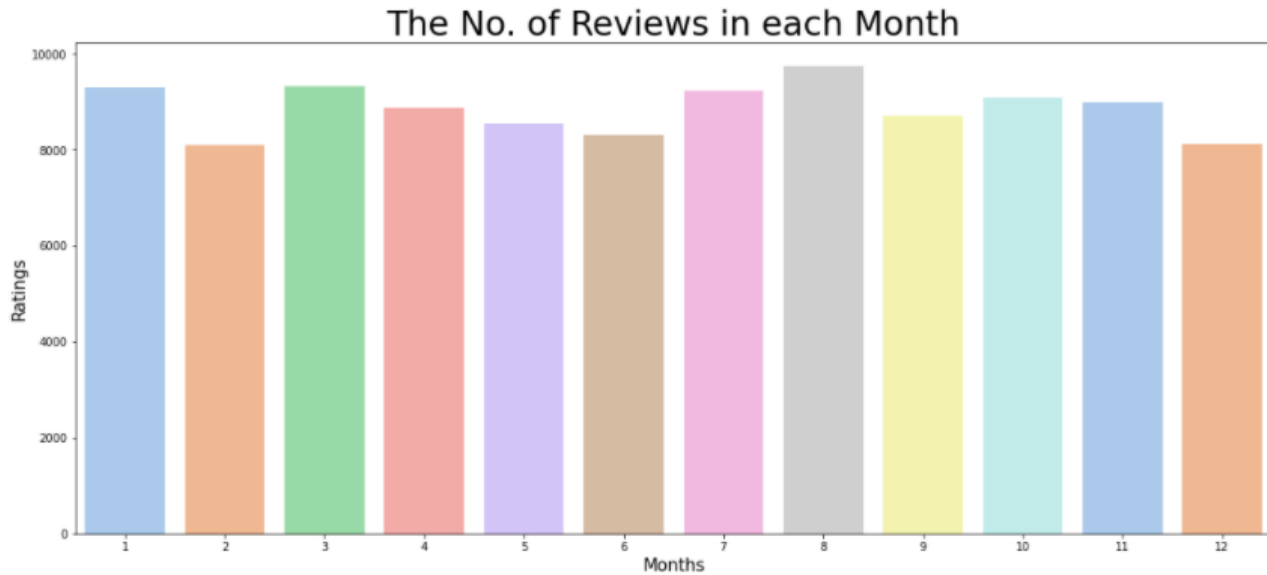
106400 rows × 7 columns

Below Bar graph tells us top 20 frequently occurring conditions in our dataset.



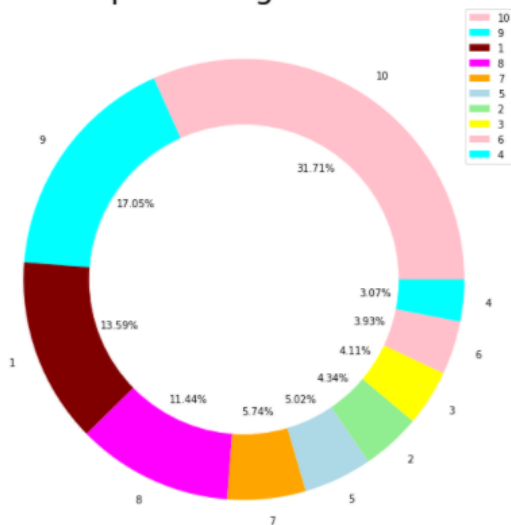
Below bar graph tells us number of reviews in each year. In 2016 there are many reviews where in 2008 there are less reviews.





Below Pie chart tells us the sharing of rating. Many users given rating 10 for many drugs along with patient condition and less rating is 4 given by people who uses drugs basing on there condition.

A Pie Chart Representing the Share of Ratings



Data Pre-Processing:

Before finding the sentiment of each review we need to pre-process our review columns using following steps

- Tokenization
- Stemming
- Lemmatization
- Stop Words Removal
- Rejoining Reviews

```
stemmer=PorterStemmer()
lemmatizer = WordNetLemmatizer()
stop_words=stopwords.words("english")
sentence_filter_stemming=[]
sentence_filter_Lemmatization=[]
final_sentence_stemming=[]
final_sentence_lemmatize=[]
#Removing noise, such as special characters and punctuations
df['Cleaned_Reviews'] = df['review'].str.replace(r"\W", " ").str.strip()#To remove special characters and punctuations
df['Cleaned_Reviews']= df['Cleaned_Reviews'].str.replace(r'\d+',"")#To remove Numbers
for a in df['Cleaned_Reviews']:
    splitting_words=word_tokenize(a)
    for b in splitting_words:
        if b not in stop_words:#Removing Stop Words
            lemmatization_words=lemmatizer.lemmatize(b.lower())#Lemmatization
            sentence_filter_Lemmatization.append(lemmatization_words)
    final_sentence_lemmatize.append(' '.join(sentence_filter_Lemmatization))
    sentence_filter_Lemmatization.clear()
df['Lemmatization']=pd.DataFrame(final_sentence_lemmatize)
df
```

Cleaned Dataset:

	uniqueID	drugName	condition	review	rating	date	usefulCount	Year	month	day	Cleaned_Reviews	Lemmatization
0	163740	Mirtazapine	Depression	"I've tried a few antidepressants over th...	10	2012-02-28	22	2012	2	28	I ve tried a few antidepressants over the ye...	i tried antidepressant year citalopram fluoxet...
1	206473	Mesalamine	Crohn's Disease, Maintenance	"My son has Crohn's disease and has done ...	8	2009-05-17	17	2009	5	17	My son has Crohn s disease and has done very...	my son crohn disease done well asacol he compl...
2	159672	Bactrim	Urinary Tract Infection	"Quick reduction of symptoms"	9	2017-09-29	3	2017	9	29	Quick reduction of symptoms	quick reduction symptom
3	39293	Contrave	Weight Loss	"Contrave combines drugs that were used for al...	9	2017-03-05	35	2017	3	5	Contrave combines drugs that were used for alc...	contrave combine drug used alcohol smoking opi...
4	97768	Cyclafem 1 / 35	Birth Control	"I have been on this birth control for one cyc...	9	2015-10-22	4	2015	10	22	I have been on this birth control for one cycl...	i birth control one cycle after reading review...
...

Now, we will find sentiments for each review with the help of Text blob package in python.

```
def sentiment(review):
    # Sentiment polarity of the reviews
    pol = []
    for i in review:
        analysis = TextBlob(i)
        pol.append(analysis.sentiment.polarity)
    return pol

df['sentiment'] = sentiment(df["Cleaned_Reviews"])
```

```
df.loc[(df['sentiment_final'] == -10) | (df['sentiment_final'] == -9) , 'updated_sentiment'] = 1
df.loc[(df['sentiment_final'] == -8) | (df['sentiment_final'] == -7) , 'updated_sentiment'] = 2
df.loc[(df['sentiment_final'] == -6) | (df['sentiment_final'] == -5) , 'updated_sentiment'] = 3
df.loc[(df['sentiment_final'] == -4) | (df['sentiment_final'] == -3) , 'updated_sentiment'] = 4
df.loc[(df['sentiment_final'] == -2) | (df['sentiment_final'] == -1) | (df['sentiment_final'] == 0) , 'updated_sentiment'] = 5
df.loc[(df['sentiment_final'] == 0) | (df['sentiment_final'] == 1) | (df['sentiment_final'] == 2) , 'updated_sentiment'] = 6
df.loc[(df['sentiment_final'] == 3) | (df['sentiment_final'] == 4) , 'updated_sentiment'] = 7
df.loc[(df['sentiment_final'] == 5) | (df['sentiment_final'] == 6) , 'updated_sentiment'] = 8
df.loc[(df['sentiment_final'] == 7) | (df['sentiment_final'] == 8) , 'updated_sentiment'] = 9
df.loc[(df['sentiment_final'] == 9) | (df['sentiment_final'] == 10) , 'updated_sentiment'] = 10
```

```
df = df.rename(columns={'updated_sentiment':'Sentiment'})
df = df.drop(columns=['sentiment'])
```

With the help of above code we will calculate final sentiment for each review.

Sample Sentiment Dataset:

Lemmatization	sentiment_final	Sentiment
i tried antidepressant year citalopram fluoxet...	0	6.0
my son crohn disease done well asacol he compl...	6	8.0
quick reduction symptom	4	7.0
contrave combine drug used alcohol smoking opi...	2	6.0
i birth control one cycle after reading review...	3	7.0
...
i androgel two month i say best i felt last ye...	1	6.0
for year i i ravenous hunger low blood sugar c...	2	6.0
so much better oxycontin osteoarthritis s i jo...	0	6.0
i taking lexapro year seems help lot social an...	0	6.0

Hypothesis:

Null Hypothesis(H0): There are many drugs for each condition.

Alternative Hypothesis(H1): There are less drugs for each condition.

Research Design and Methodology:

Collected medical data online and I am going to clean and prepare my data and then I will analyze my data then after I will design my recommendation system using NLP methods and Recommendation Techniques and then I am going to predict medicines for users and then after I am going to give Medicine overview based on user choice in that overview user can able can make decision like whether they can use this drug or not and then I will create a small web application and integrate with my Recommendation model.

Data Analytics:

Once all sentiment all calculated for each review then we will recommend medicines to users based on the user input.

Once we get input from user then based on condition we will check all the drugs that are suitable for that condition then based on Sentiment and useful count we will rank the drug and then we will recommend medicine to users.

```
input_condition = 'Birth Control'
cond_filtered = df[df['condition'] == input_condition]
cond_filtered = cond_filtered.groupby(['condition', 'drugName']).agg('mean').sort_values(['Sentiment'], ascending=False)
cond_filtered=cond_filtered.reset_index()
```

```
cond_filtered["drugName"].head(5)
```

```
0      Zovia 1 / 50
1      Altavera
2      Amethyst
3      Low-Ogestrel-21
4      Ortho Cyclen
Name: drugName, dtype: object
```

In my recommendation system users can able to see review of each drug by sentiments of each drug.

I have considered NRC-Emotion-Lexicon-Worldlevel text file and in this file, we can see sentiments of all the words.

In the below dataset we can see the reactions of each word and with the help of this we can able to find the reactions of the users who already used any drug.

emotion	word	anger	anticipation	disgust	fear	joy	negative	positive	sadness	surprise	trust
0	aback	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	abacus	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
2	abandon	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0
3	abandoned	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0
4	abandonment	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0
...
5429	germinate	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5430	germination	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5431	gestation	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5432	gesture	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5433	ghastly	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5434 rows × 11 columns

```
emotions = emotions_words.columns.drop('word')
emo_df = pd.DataFrame(0, index=new_df.index, columns=emotions)
i=0
for a in df_sample['Lemmatization']:
    splitting_words=word_tokenize(a)
    for word in splitting_words:
        emo_score = emotions_words[emotions_words.word == word]
        if not emo_score.empty:
            for emotion in list(emotions):
                emo_df.at[i, emotion] += emo_score[emotion]
    i=i+1
```

```
[61] new_df = pd.concat([new_df, emo_df], axis=1)
```


From the below dataset we can see emotions of each review.

Based on the emotions we can suggest patients whether they can prefer drug or not, because sometimes there are multiple drugs for one condition and side effects may differ for each drug, so basing on emotions of each drug patients can decide which drug is suitable for their body.

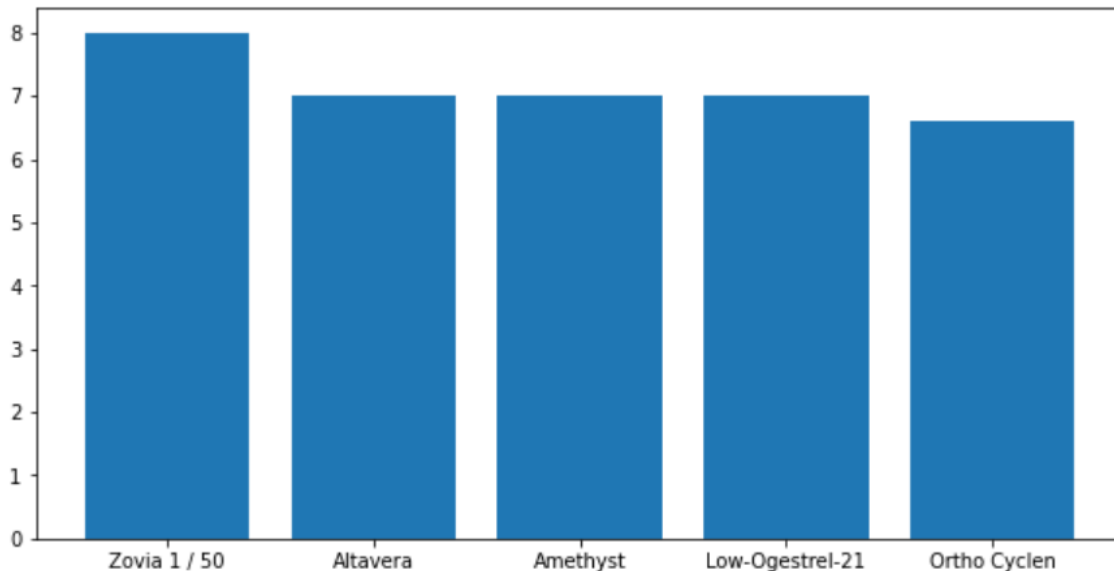
uniqueID	drugName	condition	review	rating	date	usefulCount	Year	month	day	Cleaned_Reviews	Lemmaization	sentiment_final	Sentiment	anger	anticipation	disgust	fear	joy	negative	positive	sadness	surprise	trust	
0	19262	Pregabalin	fibromyalgia	"Two attempts at treatment for fibromyalgia vi...	1	2016-11-03	33	2016	11	3	Two attempts at treatment for fibromyalgia vi...	have serquel week depression i take bed still...	-2	5.0	2	2	1	2	2	6	3	4	1	2
1	125135	Dulcolax	Constipation	"I'd been constipated for a few days and ...	10	2016-12-29	8	2016	12	29	I d been constipated for a few days and deci...	i literally taking luteru three day i hate it...	0	6.0	6	4	5	7	1	12	2	9	1	2
2	51281	Azithromycin	Chlamydia Infection	"Took the treatment this morning 4 pills at 10...	7	2015-10-05	14	2015	10	5	Took the treatment this morning pills at pl...	my primary care physician put oxycontin year h...	1	6.0	0	4	1	8	2	5	7	9	3	4
3	225155	Bupropion	Anxiety	"Just a quick note--I actually found it helpf...	8	2015-12-29	67	2015	12	29	Just a quick note I actually found it helpfu...	i neiglanon year i love the first month i unco...	3	7.0	0	3	1	1	2	3	2	1	1	1
4	52645	Canasa	Ulcerative Proctitis	"When I first started taking it it did't...	10	2016-12-30	4	2016	12	30	When I first started taking it it didn't hav...	i efflor mg daily since september mainly rei...	2	6.0	2	2	1	1	2	9	4	3	1	2
...
4996	128946	Metoprolamide	Migraine	"I went to the ER for extreme neck and head pa...	1	2017-02-17	17	2017	2	17	I went to the ER for extreme neck and head pa...	recently started treatment irritable bowel syn...	1	6.0	1	2	0	0	0	3	2	0	0	2
4996	184922	Canagliflozin	Diabetes, Type 2	"Been on Invokana for a week now. My blood sug...	8	2013-06-27	47	2013	6	27	Been on Invokana for a week now. My blood suga...	i went celera three year getting depressed anx...	2	6.0	4	4	3	8	0	12	3	8	1	1
4997	28573	Sprintec	Birth Control	"I have been on it for 4 months now. It was ok...	6	2013-06-17	12	2013	6	17	I have been on it for months now it was okay...	i well medication but i longer afford the gene...	2	6.0	0	0	0	1	0	0	1	0	0	0
4998	31496	Miconazole	Vaginal Yeast Infection	"Inserted miconazole T and with in 15 minutes ...	1	2016-06-21	7	2016	6	21	Inserted miconazole and with in minutes the ...	the excruciating pain i experiencing period ve...	2	6.0	0	0	1	3	0	3	0	2	0	0
4999	85156	Ethinyl estradiol / norgestimate	Birth Control	"This birth control has worked 100% with prove...	3	2015-11-07	11	2015	11	7	This birth control has worked with preventin...	pretty much ruined skin ive using episoo month...	0	6.0	2	2	2	2	2	3	3	1	0	4
5000 rows x 24 columns																								

Data Visualization and Result Reports:

Drug Recommendation:

```
plt.figure(figsize=(10,5))
plt.bar(cond_filtered["drugName"].head(5),cond_filtered["Sentiment"].head(5))
```

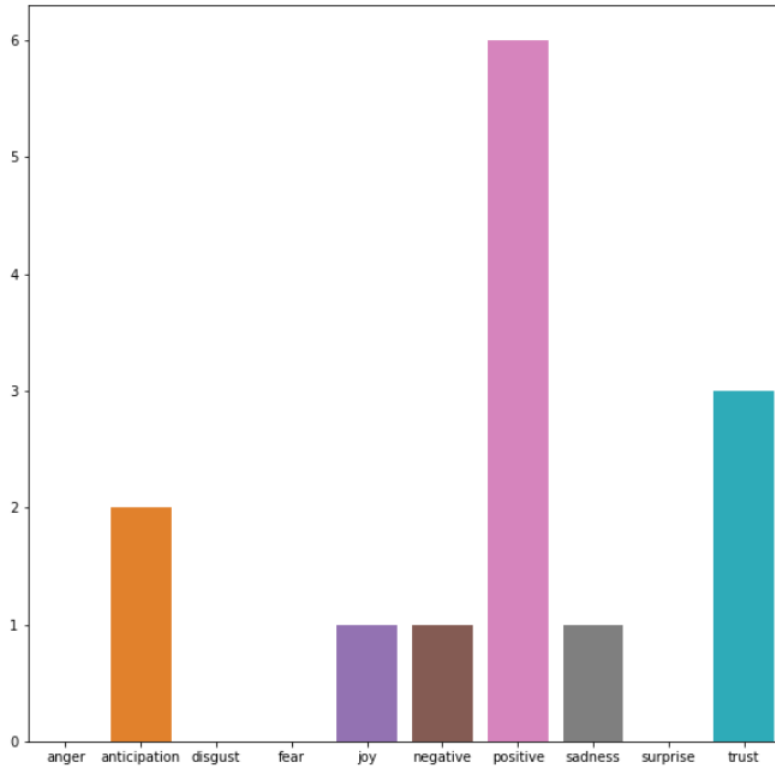
<BarContainer object of 5 artists>



In the above bar graph we have recommended top 5 drugs for Birth Control condition and patient can see what are the emotions or any side effect of any particular drug.

About Drug:

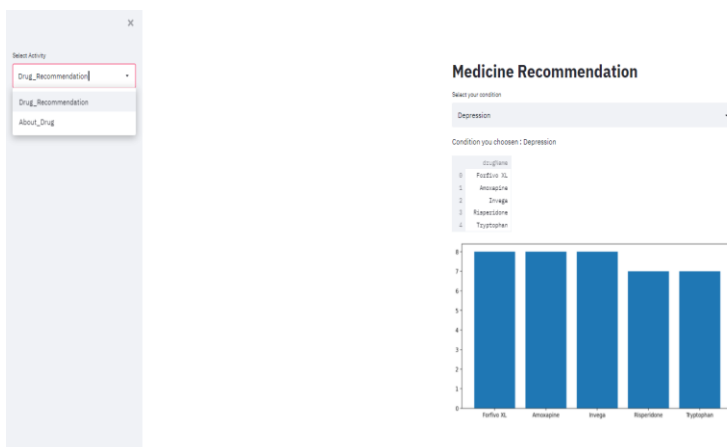
```
df=pd.read_csv("/content/Sentiment_Emotions.csv")
df=df[df["drugName"]=="Magnesium sulfate / potassium sulfate / sodium sulfate"].sort_values(by="usefulCount", ascending=False).iloc[0,15:25]
```



So from above bar graph we can see drug emotions.

Web Application:

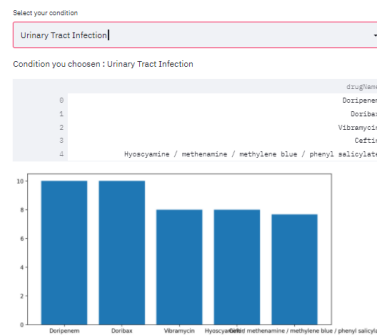
Created a Simple Recommendation web application using streamlit module in python.



Select Activity

Drug_Recommendation

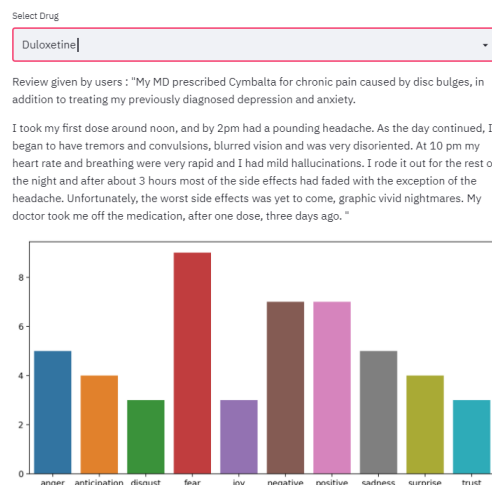
Medicine Recommendation



Select Activity

About_Drug

Medicine Recommendation



Conclusion:

From Machine Learning Algorithm we can develop recommender systems that can give a medical recommendation with excellent efficiency and accuracy based on diagnosis and symptoms. The results using NLP methods and Recommendation will help in knowing the drugs based on condition.

Bibliography:

References

[1] Recommendation of Doctors and Medicines Using Review Mining," The FDA Safety Information and Adverse Event Reporting Program, 18 Apr. 2018, pp. 1-2.

- [2] RPubS - Project 5: Medical Recommender System. (2019b, May 12). Rajwant Mishra. https://rpubs.com/Rajwantmishra/project5_final/, 2012.
- [3] Kordík, P. (2019, December 15). Machine Learning for Recommender systems — Part 1 (algorithms, evaluation, and cold start). Medium. <https://medium.com/recombee-blog/machine-learning-for-recommender-systems-part-1-algorithms-evaluation-and-cold-start-6f696683d0ed>.
- [4] Bansal, M. (2019, October 18). Decoding the output of a hybrid recommendation system. Medium. <https://towardsdatascience.com/decoding-the-output-of-a-hybrid-recommendation-system-2007d85547bf>
- [5] Cates, J. (2019, October 4). How to Build a Recommender Engine for Medical Research Papers. Medium. <https://medium.com/@topspinj/how-to-build-a-recommender-engine-for-medical-research-papers-a9dc1bad12f2>