



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: 1 Month of publication: January 2020

DOI: <http://doi.org/10.22214/ijraset.2020.1068>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Association Rule Mining of Inactive Ingredients in Drugs

Mohammed Azam Sayeed¹, Sakeena Fiza², Noor Ayesha³, Mohammed Akram Sayeed⁴

¹Department of Computer sciences, SET Jain University Bangalore -562112

²UG Student at UC Berkeley

Abstract: *This paper presents a proposed methodology to extract strong association rules by frequently occurring inactive ingredient combinations using the concept of Association Rule Mining which will help identify the inactive ingredient contributing the most to popularity levels and rating of the medicines available in market, assisting in curating better marketable and effective medicines and also assist in reducing unnecessary inactive ingredients to be excluded in formulation.*

Keywords: *Association Rule Mining, Apriori Principle, Inactive ingredients, Active ingredients, Itemset, Antecedent, Consequent, support, confidence, lift, generic name, Medications for fever, combinational ailments, inflammation, acetaminophen systemic, ibuprofen systemic, aspirin, Prednisolone, pruning, parameter tuning, medicine formulation.*

I. INACTIVE INGREDIENTS IN DRUGS

Frank Wells writes in Colourings And Preservatives In Drugs published in the British Medical Journal, "It was in March 1988 that the Association of the British Pharmaceutical Industry published guidelines on the disclosure of inactive ingredients. It made strong recommendations that there should be full qualitative disclosure for injectable, inhalant, and topical drugs, including ophthalmic preparations, and there should be a qualitative disclosure of particular inactive ingredients that had been identified as a possible hazard for certain patients for oral and other formulations." (1099). However, the labeling of these ingredients that pose a prominent threat to the consumer remains voluntary. The presence of inactive ingredients in pharmaceutical products such as - binders, fillers, dyes, and preservatives - contribute to the improvement of the appearance, bioavailability, stability, and palatability of the product. These pharmaceutical adjuvants are known to be non-medicated part of the drug, they remain inert and do not interfere with the intended action of the prescribed medication. Clinical Savvy: How Safe Are Generic Drugs? published in The American Journal of Nursing describes inactive ingredients as, "nothing more than a delivery system for the active ingredient." (Birdsall & Uretsky 431). But these ingredients - that constitute the majority of the mass and volume of oral and parenteral drugs - are not as inactive as they are made out to be. In the database run by the National Library of Medicine, data about inactive ingredients shows that active ingredients only account for just over a quarter (29 percent) of the weight of an oral pill, whereas the inactive ingredients constitute the remaining 71 percent. Once an inactive ingredient has appeared in an approved drug, it is added to the inactive ingredient database which contains information on inactive ingredients present in FDA-approved drug products. This provides leeway in new drug development, as the inactive ingredient is not considered new and does not require an extensive review. Currently, an oral pill contains about eight inactive ingredients on average and can have up to thirty-five. Drug companies have access to a surplus of these ingredients such as but not limited to - benzoic acid and its derivatives, ethanol, gluten, and peanut oil - whilst manufacturing pills. Some of these ingredients are allergens, even though they are tested for toxicity and do not affect a majority of the population, they have shown to adversely react in some patients. Studies have found that a majority of the pills contain at least one of these ingredients that prove to be problematic for patients with certain food intolerances, such as gluten and sugar. In this paper, the adverse reactions triggered by an inactive ingredient are explored further to narrow down and develop a mechanism to avoid the ingredients that are known to be allergens and focus on the inactive ingredients that are necessary, in combination with the active pharmaceutical ingredients for the specific pharmaceutical formulation.

In certain situations, inactive ingredients can become they do happen and can have a very real impact on the health of those taking the medication. Despite some doctors occasionally seeing some patients have adverse side effects from inactive ingredients, a study had never been done regarding it. That is, until recently when a team of researchers decided to look into it to see just how dangerous these additives could be. They discovered that the average pill contains around 25 percent of the active ingredient while the rest is the inactive additives. This isn't very surprising because many medications are active at very small doses. Without additives like fillers, these drugs would be minuscule powders. More interestingly is that some of the most common side effects that were experienced involved things like food allergies, lactose intolerance, and celiac disease. Lactose in specific was found in 45 percent

of the pills that they checked. For those with allergies, the inactive ingredients in medications can be a pretty big minefield. Tons of different potential allergens are included as additives in medications and this is an incredibly common practice for all pharmaceutical companies. Several cases have been reported where inactive ingredients have become potential allergens for patients with allergies. It is always a good practice to be cautious and to double-check what kinds of inactive ingredients a medication contains for Health workers and patients alike. The researcher advises it is better to always have lesser inactive ingredients entering the human body by medicines consumption as it could have side effects and potential to become active on certain triggers and could be absorbed by the body in its potential active states.

The paper will work on the study of the frequently occurring inactive ingredients grouped by several important use cases, finding strong association rules by frequently occurring ingredient combinations using the concept of Association Rule Mining Explained in the next sections. The insights from the strong rules of itemset of inactive ingredients have multiple applications such as precision medicines by reducing inactive ingredients, generate marketing insights having awareness of which inactive ingredients contribute the most to medicines popularity and rating and also assist in designing better medicines formulation for impact, curated and study to reduce the inactive ingredient in Drugs.

II. INTRODUCTION OF ASSOCIATION RULE MINING

A. The intuition of Association Rule Mining:

Association Rule Mining is a rule-based Machine Learning method that is used to identify frequently occurring patterns, associations and highly correlated relationships between Entities. “Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or another information repository. An example of an association rule would be "If a customer buys a dozen eggs and bread, he is 80% likely to also purchase milk. “An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item found in the data. A consequent is an item that is found in combination with the antecedent.” [1] (Gurpreet Singh). Below is an example of an Association Rule.

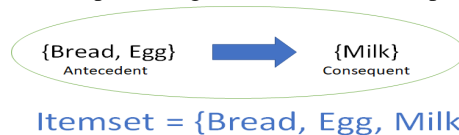


Fig. 1 Association Rule

Association Rule Mining can uncover important associations between store items to maximize cross-sales, promotional offerings, catalog management, efficient marketing and revenue improvising.

B. Measures of Association Rule Mining:

We have various metrics to find the strength of the association and filters to avoid generations of Huge Numbers of insignificant rules generated from Apriori principle Frequently occurring itemset. They are as follows –

1) *Support*: Support determines the proportion of the overall dataset which has both the antecedent and the consequent present in it.

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Total\ number\ of\ transactions}$$

Fig. 2 Mathematical representation of Support

Value of Support helps us to identify if the discovered rules are worth considering for analysis or not

Ex: Consider we have occurrences of itemset {X, Y} of only 50 times out of a total of 10,000 transactions i.e. support = 0.005. If an itemset happens to have very low support, we do not have enough information on the relationship between its items and hence no conclusions can be drawn from such a rule. Support is a measure used in the Apriori principle to generate frequent itemset, will be discussed in section C.

2) *Confidence*: Confidence refers to the likelihood that the item Y is also bought if Item X is bought

$$Confidence(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Transactions\ containing\ X}$$

Fig. 3 Mathematical representation of Confidence

Ex: If Confidence of $\{X\} \Rightarrow \{Y\}$ is high, that means the association is strong and if X is purchased/Present in the transaction then most likely Y is also to be present in the same transaction.

Drawback: Confidence of an association rule having very frequent/popular consequent item will result in high Confidence regardless of the frequency/popularity of Antecedent Item. We can overcome this drawback by third measure Lift

3) *Lift*: Lift refers to the likelihood of Item Y being purchased when Item X is purchased while controlling how popular Item Y is.

$$Lift(\{X\} \rightarrow \{Y\}) = \frac{(Transactions\ containing\ both\ X\ and\ Y) / (Transactions\ containing\ X)}{Fraction\ of\ transactions\ containing\ Y}$$

Fig. 4 Mathematical representation of Lift

Or can be represented as Support of Itemset over the product of Support of Antecedent and Consequent

The measure is used as below

If $Lift(X \Rightarrow Y) > 1$; Y is likely to be bought if X is bought

If $Lift(X \Rightarrow Y) < 1$; Y is unlikely to be bought if X is bought

C. Apriori Algorithm

A brute force approach to find frequent itemsets is to form all possible itemsets and check the support value of each of these. Apriori principle helps in making this search efficient.

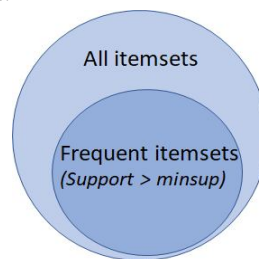


Fig. 5 Frequent Itemset using support

Apriori principle says that any subset of a frequent itemset must also be frequent.

Ex: a transaction containing {wine, chips, bread} also contains {wine, bread}. So, according to the principle of Apriori, if {wine, chips, bread} is frequent, then {wine, bread} must also be frequent.

[1] The Apriori Algorithm is an influential algorithm for mining frequent itemsets for boolean association rules.

Key Concepts:

- 1) Frequent Itemsets: The sets of an item that has minimum support (denoted by L_i for i th-Itemset).
- 2) Apriori Property: Any subset of a frequent itemset must be frequent.
- 3) Join Operation: To find L_k , a set of candidate k -itemsets is generated by joining L_{k-1} with itself.
- 4) Find the frequent itemsets: the sets of items that have minimum support – A subset of a frequent itemset must also be a frequent itemset
- a) if $\{AB\}$ is a frequent itemset, both $\{A\}$ and $\{B\}$ should be a frequent itemset
- b) Iteratively find frequent itemsets with cardinality from 1 to k (k -itemset)
- 5) Use frequent itemsets to generate association rules.

The Apriori Algorithm : Pseudo code

```

• Join Step:  $C_k$  is generated by joining  $L_{k-1}$  with itself
• Prune Step: Any  $(k-1)$ -itemset that is not frequent cannot be a subset of a frequent  $k$ -itemset
• Pseudo-code:
 $C_k$ : Candidate itemset of size  $k$ 
 $L_k$ : frequent itemset of size  $k$ 

 $L_1 = \{ \text{frequent items} \};$ 
for ( $k = 1; L_k \neq \emptyset; k++$ ) do begin
 $C_{k+1} =$  candidates generated from  $L_k$ ;
for each transaction  $t$  in database do
increment the count of all candidates in  $C_{k+1}$ 
that are contained in  $t$ 
 $L_{k+1} =$  candidates in  $C_{k+1}$  with min_support
end
return  $\cup_k L_k$ 

```

Fig. 6 Apriori Algorithm

In the above pseudo-code, we are using the Apriori principle which allows us to prune all the supersets of an itemset that does not satisfy the minimum threshold condition for support. Generate all frequent itemsets (support \geq minsup) having only one item. Next, generate itemsets of length 2 as all possible combinations of the above itemsets. Then, prune the ones for which support value fell below minsup. Now generate itemsets of length 3 as all possible combinations of length 2 itemsets (that remained after pruning) and perform the same check on support value.

We keep increasing the length of itemsets by one like this and check for the threshold at each step.

Then from step 5. Generate association rules from frequent itemsets (support \geq minsup) can be pruned based on threshold value specified by the confidence and lift threshold values. Pruning is depicted in Fig.

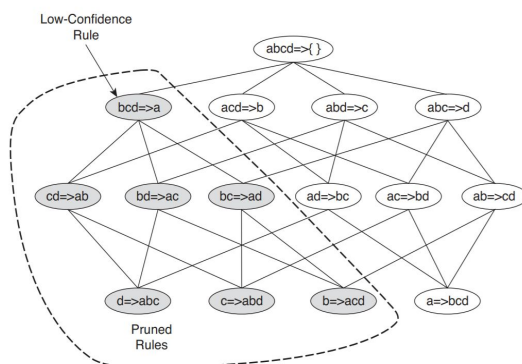


Fig. 7 Pruning on rules using Confidence and Lift

III. PROPOSED METHODOLOGY

The proposed Methodology can be briefly depicted as Fig 8. We are using python jupyter notebook interface for analyzing inactive ingredients and apriori API from the apyori package. Can be installed in Anaconda Prompt the command \$pip install apyori

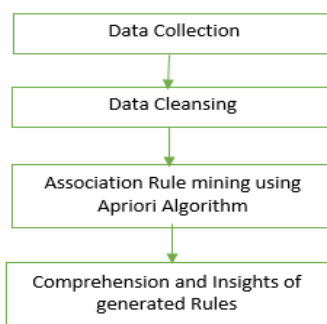


Fig. 8 Proposed Methodology

A. Data Collection

We have collect data on various medicines based on generic drug name which is prescribed for specific ailment and based on popularity/rating provided in widely accepted Medically approved US site for drug details <https://www.drugs.com/> and various brand names having generic drug name as active ingredient, the drug’s composition containing the inactive ingredients data were gathered from various sites primarily from <https://dailymed.nlm.nih.gov/dailymed/>. The data is stored in CSV format.

B. Data Cleansing

The data collected can have few inactive ingredients with the same chemical composition but may have been referred by different common names as data was collected from various resources such as “Pregelatinized starches” is the same as “Starch”. So any mention of “Starch” was considered as “Pregelatinized starches”. However, very few such occurrences were recorded. Also, this step removes spaces between commas in the CSV file, conversion to lower cases and avoid duplication of inactive ingredients to reduce effort in Step C. Also we have removed any mg concentration details as our paper doesn't focus on strength of ingredients but rather on the presence of various inactive ingredients formulated by pharmaceutical manufacturing templates.

Steps A. and B. have been depicted clearly in Table 1-4.

C. Association Rule Mining using Apriori Algorithm

- 1) *Read data cleansed CSV file:* The data saved in the CSV file is read and processed using the Pandas package in python.

```
import pandas as pd
acetaminophen_data = pd.read_csv('acetaminophen.csv',header=None)
```

Fig. 9 Reading CSV file

- 2) *Generate List of Records for Apriori API:* This step is necessary because apriori API from the apyori package works on records as list and not as pandas Dataframe. An example of List of record is depicted in Fig 10.

We have also Filled None to Nan Values and optional steps of removing spaces between commas and converting to the lower case can be done before converting to List of records.

```
In [81]: records=[]
         for i in range(0,s.shape[0]):
           records.append([str(s.values[i,j]) for j in range(0,s.shape[1])])
```

```
In [82]: records[0]
```

```
Out[82]: ['acetaminophen',
          'cellulose',
          'cornstarch',
          'hypromellose',
          'magnesium stearate',
          'polyethylene glycol',
          'sodium starch glycolate',
          'nan',
          'nan',
          'nan',
          'nan',
          'nan',
          'nan',
          'nan',
          'nan',
          'nan']
```

Fig. 10 List of Records for apriori API

- 3) *Generate Rules based on Apriori API and Parameter Tuning:* The list of records from the dataset is fed into the apriori API to generate a list of frequently occurring association rules. The most important pointer in this step is parameter tuning. Parameter Tuning is important otherwise apriori will generate an incredibly large number of insignificant rules for itemset based on apriori principle which is computationally expensive and analytically insignificant. Thus we use parameters such as Measures of Association rule mining for rule pruning and removing all incognizant rules. We are using min_support value as 0.4, which means occurrences of itemset containing antecedent and consequent should be at least greater than 40% of the total transaction which will be used for generation of frequent itemset

```
from apyori import apriori
association_rules =apriori(records,min_support=0.4,min_confidence=0.6,min_lift=1.5)
association_results=list(association_rules)
print(len(association_results))
```

Fig. 11 Apriori API with parameter tuning

We specify other measures such as min_confidence =0.6 and min_lift = 1.5 this will ensure that insignificant rules will not be generated from the apriori algorithm and will be used in rule pruning step

4) Saving on Rule Itemset and Measures Values in xls Format for the Next Step

Out[88]:

	Items	Antecedent	Consequent	Support	Confidence	Lift
0	{sodium starch glycolate, cellulose}	{cellulose}	{sodium starch glycolate}	0.5	1.000000	1.666667
2	{sodium starch glycolate, povidone}	{povidone}	{sodium starch glycolate}	0.4	1.000000	1.666667
4	{sodium starch glycolate, cellulose, acetamino...}	{cellulose}	{sodium starch glycolate, acetaminophen}	0.5	1.000000	1.666667
6	{sodium starch glycolate, cellulose, acetamino...}	{cellulose, acetaminophen}	{sodium starch glycolate}	0.5	1.000000	1.666667
8	{sodium starch glycolate, acetaminophen, povid...}	{povidone}	{sodium starch glycolate, acetaminophen}	0.4	1.000000	1.666667
10	{sodium starch glycolate, acetaminophen, povid...}	{acetaminophen, povidone}	{sodium starch glycolate}	0.4	1.000000	1.666667
14	{sodium starch glycolate, cellulose, nan}	{cellulose, nan}	{sodium starch glycolate}	0.4	1.000000	1.666667
20	{sodium starch glycolate, cellulose, acetamino...}	{cellulose, nan}	{sodium starch glycolate, acetaminophen}	0.4	1.000000	1.666667
22	{sodium starch glycolate, cellulose, acetamino...}	{cellulose, acetaminophen, nan}	{sodium starch glycolate}	0.4	1.000000	1.666667
1	{sodium starch glycolate, cellulose}	{sodium starch glycolate}	{cellulose}	0.5	0.833333	1.666667
5	{sodium starch glycolate, cellulose, acetamino...}	{sodium starch glycolate}	{cellulose, acetaminophen}	0.5	0.833333	1.666667
7	{sodium starch glycolate, cellulose, acetamino...}	{sodium starch glycolate, acetaminophen}	{cellulose}	0.5	0.833333	1.666667
3	{sodium starch glycolate, povidone}	{sodium starch glycolate}	{povidone}	0.4	0.666667	1.666667

Fig. 12 Save Results in xls for assessments

D. Comprehension and Insights of generated Rules

The step will involve assessment and studying of strong association rules from the frequently occurring inactive ingredients. We will understand which all frequently occurring inactive ingredients combinations are present in itemset along with the active ingredient. This will help in understanding which inactive ingredients in combination with active ingredient would be preferable to include for new drug formulations for better marketing, custom medication, etc and also figure out which of them are less occurring inactive ingredients that can be considered to be avoided or excluded in drug preparations.

To understand this step clearly, we have attempted to apply the proposed methodology for a few significant business use cases in the pharmaceutical domain and understand step D in detail.

IV. USE CASE STUDY

A. Use Case: Highest Rated Drugs For Specific Ailment (Considering Fever)

Based on Medications for fever we referred <https://www.drugs.com/condition/fever.html> site and found the most popular and high rated Active ingredient, generic drug class. For our research, we selected the Generic name: acetaminophen systemic drug.

Description of Acetaminophen: Acetaminophen is a pain reliever and a fever reducer. Acetaminophen is used to treat mild to moderate and pain, to treat moderate to severe pain in conjunction with opiates, or to reduce fever. Common conditions that acetaminophen treats include headache, muscle aches, arthritis, backache, toothaches, colds, and fevers.

Corresponding US brand name drugs were collected having acetaminophen as its active ingredient along with its inactive ingredients from various resources such as <https://dailymed.nlm.nih.gov/dailymed/>. Refer to the below table for a few of the collected data on medicines.

TABLE I. BRAND Names for Generic name: acetaminophen systemic

Brand Name	Compostion (csv format)
tylenol	acetaminophen,cellulose,cornstarch,hypromellose,magnesium stearate,polyethylene glycol,sodium starch glycolate,ferric oxide
children's tylenol	acetaminophen,citric acid,glycerin,high fructose corn syrup,cellulose,sodium starch glycolate,purified water,sodium benzoate,sorbitol solution,sucralose,xanthan gum
acephen	acetaminophen,glyceryl stearate,hydrogenated vegetable oil,polyethylene glycol,magnesium stearate,sorbitan monooleate
mapap	acetaminophen,hypromellose,polyethylene glycol,povidone,pregelatinized starch,sodium starch glycolate,stearic acid
childrens qpap	acetaminophen,butylparaben,citric acid,glycerin,high fructose corn syrup,hypromellose,purified water,sodium benzoate,sorbitol solution,xanthan gum
ofirmev	acetaminophen,mannitol,cysteine hydrochloride,sodium phosphate,dibasic,sodium hydroxide,hydrochloric acid
atenolol	acetaminophen,citric acid,sodium starch glycolate,magnesium stearate,silicon dioxide,povidone,cellulose,microcrystalline
cetafen	acetaminophen,corn syrup,starch,microcrystalline,cellulose,povidone,sodium starch glycolate,stearic acid
Equaline Pain Relief	acetaminophen,crospovidone,edetate disodium,gelatin,glycerin,hypromellose,magnesium stearate,microcrystalline,cellulose,polyethylene glycol,povidone,pregelatinized starch,sodium starch glycolate,stearic acid,titanium dioxide
Feverall	acetaminophen,glycerol monostearate,hydrogenated vegetable oil,polyoxyethylene stearate,polysorbate 80

On performing steps C. from proposed Methodolgy , we can generate rules from frequent itemset such as below

{'cellulose'}=>{'sodium starch glycolate', 'acetaminophen'}
 {'cellulose', 'acetaminophen'}=>{'sodium starch glycolate'} etc

1) *Results:* We extracted combinational itemset containing active ingredient Acetaminophen as follows {acetaminophen,cellulose,povidonce,Sodium starch glycolate}. This means in popular and highly rated active ingredient acetaminophen for medication for fever, most of the brand names have a combination of cellulose, providence and sodium start glycolate frequently occurring.

Thus better formulation of drugs it is optimal to include these combinations of inactive ingredients as these combinations are quite popular and highly rated by end users also we could work on reducing the number of other inactive ingredients (such as polyethylene glycol, butylparaben, etc) as they don't seem to contribute much to the popularity and efficiency of drugs based rules generated on parameter tuning values.

Example coloring and pigments (such as ferric oxide etc) and emulsifiers (such as polysorbate 80) were found to have no impact on drug rating which is intuitive. In-depth inactive ingredient uses can be explored using online resources such as www.drug.com, Inactive ingredient uses can be found as below

Sodium starch glycolate uses - <https://www.drugs.com/inactive/sodium-starch-glycolate-128.html> , Sodium starch glycolate absorbs water rapidly, resulting in swelling which leads to the rapid disintegration of tablets and granules. Which makes sense to be included in most of the brand names.

The same Procedure can be used for any other Highest Rated Generic Drug Name for Fever as explained below

Considering Generic name: ibuprofen systemic

TABLE 2. BRAND Names for Generic name: ibuprofen systemic

Brand Name	Compostion (csv format)
Advil	ibuprofen,acesulfame potassium,caramel color,carnauba wax,colloidal silicon dioxide,copovidone,hypromellose,mannitol,medium-chain triglycerides,microcrystalline,cellulose,polyethylene glycol,propylene glycol,sodium lauryl sulfate,sucralose,titanium dioxide
Motrin	ibuprofen,acesulfame potassium,anhydrous citric acid,glycerin,polysorbate 80,pregelatinized starch,purified water,sodium benzoate,sucrose,xanthan gum
IBU	ibuprofen,carnauba wax,silicon dioxide,croscarmellose sodium,hypromelloses,magnesium stearate,cellulose,microcrystalline,polydextrose,Polyethylene Glycol,polysorbate 80,titanium dioxide
Motrin Childrens	ibuprofen,acesulfame potassium,anhydrous citric acid,glycerin,polysorbate 80,pregelatinized starch,purified water,sodium benzoate,sucrose, xanthan gum
Advil Children's	ibuprofen,acesulfame potassium,caramel color,carnauba wax,colloidal silicon dioxide,copovidone,hypromellose,mannitol,medium-chain triglycerides,microcrystalline,cellulose,polyethylene glycol,propylene glycol,sodium lauryl sulfate,sucralose,titanium dioxide
Addaprin	ibuprofen,carnauba wax,sodium starch,hypromellose,lactose,magnesium stearate,microcrystalline cellulose,polydextrose,polyethylene glycol,polyvinyl alcohol,povidone K30,silicon dioxide,sodium starch glycolate,stearic acid,talc,titanium dioxide
Motrin IB	ibuprofen,gelatin,polyethylene glycol,potassium hydroxide,purified water,sorbitan,sorbitol
Nuprin	ibuprofen,silicon dioxide,magnesium stearate,cellulose,microcrystalline,pregelatinized starch,corn,sodium starch,sodium starch ,glycolate,hypromelloses,titanium dioxide,triacetin
Proprial	ibuprofen,carnauba wax,croscarmellose sodium,hydroxypropyl methylcellulose,microcrystalline cellulose,polyethylene glycol,polysorbate 80,povidone, pregelatinized starch,silicon dioxide,sodium starch glycolate,stearic acid,titanium dioxide
Advil Liqui-Gels	ibuprofen,acesulfame potassium,caramel color,carnauba wax,colloidal silicon dioxide,copovidone,hypromellose,mannitol, medium-chain triglycerides,microcrystalline,cellulose,polyethylene glycol,propylene glycol,sodium lauryl sulfate,sucralose,titanium dioxide

2) *Results:* We extracted combinational itemset containing active ingredient as ibuprofen as follows {ibuprofen,cellulose,carnauba wax,microcrystalline,titanium dioxide,hypromellose,polyethylene glycol } for the given parameter tuning values. Thus similar conclusions can be made to identify necessary and contributing inactive ingredients for product rating and identify other insignificant inactive ingredients to help reduce/eliminate for better formulation and avoiding unnecessary chemicals to be absorbed by the patients.

This methodology works well for any Generic drug name and shows no bias result generation

B. Use Case: Lowest Rated Drugs For Specific Ailment (Considering Fever)

Based on Medications for fever we referred <https://www.drugs.com/condition/fever.html> site and found the less popular and low rated Active ingredient, generic drug class. For our research, we selected the Generic name: aspirin/caffeine systemic. We will not disclose the brand name for this case as rating/popularity can vary over time.

TABLE 3. Generic name: aspirin / caffeine systemic

Sno	Compostion (csv format)
1.	aspirin,citric acid,sodium bicarbonate,docusate sodium
2.	aspirin,docusate sodium,fumaric acid,lactose monohydrate,potassium chloride
3.	aspirin,docusate sodium,fumaric acid,lactose monohydrate,potassium chloride
4.	aspirin,docusate sodium,sodium bicarbonate
5.	aspirin,caffine,docusate sodium,fumaric acid,lactose monohydrate,potassium chloride
6.	aspirin,caffine,docusate sodium,fumaric acid,lactose monohydrate,potassium chloride,corn starch
7.	aspirin,carnauba wax,corn starch,hypromellose,cellulose,propylene glycol,shellac,titanium dioxide,triacetin,docusate sodium
8.	aspirin,caffine,corn starch,hypromellose,microcrystalline cellulose,polyethylene glycol,sodium lauryl sulfate
9.	aspirin,caffine,corn starch,hypromellose,microcrystalline cellulose,polyethylene glycol,sodium lauryl sulfate
10.	aspirin,cellulose,corn starch,hypromellose,sodium bicarbonate,sodium lauryl sulfate,talc,titanium dioxide, triacetin

1) *Results:* We extracted combinational itemset containing active ingredient aspirin as follows { aspirin,fumaric acid,docusate sodium,lactose monohydrate,potassium chloride,cornstarch } for the giving parameter tuning values. This means in low in popularity and rating active ingredient aspirin for medication for fever, brand names have a combination of fumaric acid, docusate sodium, lactose monohydrate, potassium chloride, and cornstarch frequently occurring. A thus better formulation of drugs it is advisable to exclude these combinations of inactive ingredients as they seem to contribute to the low popularity and rating of drugs based on parameter tuning values. It is quite intuitive that inactive ingredients such as lactose monohydrate maybe not preferred by lactose-intolerant patients and coloring agents like titanium dioxide have no significant importance.

C. *Use Case: Combinational Drugs for Inflammation (Allergies, arthritis, acne inflammations)*

The same exercise can be repeated for Drugs prescribed for combinational ailments for the same patients, Prednisolone was selected for the drug prescribed for Inflammation caused by allergies, arthritis, and acne for the same patient

TABLE 4. BRAND Names for Generic name: ibuprofen systemic

Brand Name	Compostion (csv format)
Flo-Pred	prednisolone,butylparaben,carbomer homopolymer type b,edetate disodium,glycerin,poloxamer 188,propylene glycol,water,sodium hydroxide,sucralose
Millipred	prednisolone,anhydrous lactose,silicon dioxide,crospovidone,docusate sodium,magnesium stearate,sodium benzoate
Orapred	prednisolone,citric acid monohydrate,silicon dioxide,crospovidone,hypromelloses,magnesium stearate,mannitol,cellulose,microcrystalline,sodium bicarbonate,sucralose,sucrose,methacrylic acid
Orapred ODT	prednisolone,citric acid monohydrate,silicon dioxide,crospovidone,hypromelloses,magnesium stearate,mannitol,cellulose,microcrystalline,sodium bicarbonate,sucralose,sucrose,methacrylic acid
Pediapred	prednisolone,sodium phosphate,dibasic,edetate disodium,methylparaben,water,monopotassium phosphite,sorbitol
Deltasone	prednisolone,citric acid monohydrate,magnesium stearate,microcrystalline,cellulose,starch,corn,sodium glycolate
Rayos	prednisolone,citric acid monohydrate,povidone,croscarmellose sodium,silicon dioxide ,magnesium stearate,dibasic calcium phosphate dihydrate,glyceryl dibehenate,water
Predacort 50	prednisolone,citric acid monohydrate,silicon dioxide,crospovidone,hypromelloses,magnesium stearate,mannitol,cellulose,microcrystalline,sodium bicarbonate,sucralose,sucrose,methacrylic acid
Veripred 20	prednisolone,anhydrous lactose,silicon dioxide,crospovidone,docusate sodium,magnesium stearate,sodium benzoate

1) *Results:* We extracted combinational itemset containing active ingredient aspirin as follows {prednisolone, cellulose, citric acid monohydrate, magnesium stearate, microcrystalline, crospovidone, silicon dioxide} for the given parameter tuning values. This means for patients suffering from combinational ailments like inflammation caused by allergy, arthritis and severe acne can have medicines formulation having cellulose, citric acid monohydrate, magnesium stearate, microcrystalline, crospovidone, silicon dioxide as inactive ingredients. Other inactive ingredients can be reduced/minimized based on in-depth research of health care professionals.

The objective in this Use case was to find frequently occurring itemset and attempt to reduce the number of inactive ingredients used in drug formulation.

V. APPLICATIONS OF METHODOLOGY

The proposed system is capable of generating important insights about the combinational set of inactive ingredients with the active ingredient based on popularity and rating. This business value can be used in better medicine formulation to understand which inactive ingredients contribute the most to popularity and rating in the market, the foundation for research in pharmaceuticals to reduce the insignificant inactive ingredients thereby reducing potential harmful chemicals being absorbed by the patient's body and give a boost to precision medicines research.

REFERENCES

- [1] Gurpreet Singh, Sonia Jassi. "A Review Paper: A Comparative Analysis on Association Rule Mining Algorithms" *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-6 Issue-2, May 2017
- [2] Wells, Frank. "Colourings And Preservatives In Drugs." *BMJ: British Medical Journal*, vol. 299, no. 6707, 1989, pp. 1099–1099. JSTOR, www.jstor.org/stable/29705832. Accessed 16 Jan. 2020.
- [3] Birdsall, Carole, and Samuel Uretsky. "Clinical Savvy: How Safe Are Generic Drugs?" *The American Journal of Nursing*, vol. 87, no. 4, 1987, pp. 431–432. JSTOR, www.jstor.org/stable/3470427. Accessed 16 Jan. 2020.
- [4] Bharati M. Ramageri, "DATA MINING TECHNIQUES AND APPLICATIONS," *Indian Journal of Computer Science and Engineering* Vol. 1 No. 4 301-305
- [5] YanxiLiu, "Study on Application of Apriori Algorithm in Data Mining", *IEEE Second International Conference on Computer Modeling and Simulation*, 978-0-7695-3941-6/10, 2010.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)