Apriori based Medicine Recommendation System

Indrashis Mitra[0000-0002-6679-0602], Souvik Karmakar , Kananbala Ray [0000-0002-2796-995X] and T. Kar[0000-0001-9020-3626]

KIIT Deemed to be University, Bhubaneswar, Odisha 750124, India
1807274@kiit.ac.in, 1807228@kiit.ac.in, kbrayfet@kiit.ac.in
tkarfet@kiit.ac.in

Abstract. Due to a lack of preparedness, the COVID-19 epidemic wreaked havoc on the healthcare system. COVID-19 has an increased risk of causing catastrophic results in some vulnerable groups, such as the old, fragile, or those with several chronic illnesses. The scarcity of medications also played a significant role in the large number of deaths we witnessed.

Essential pharmaceutical shortages have been recorded in high, middle-, and low-income nations alike. They are costly to handle for health systems, resulting in increased expenses for drug replenishment and using substantial staff effort. Patients' health is at danger due to non-treatment, under-treatment, and possible medication mistakes caused by attempts to substitute missing drugs.

The importance of data collection, integration, processing, and reporting of underlying knowledge, as well as how this knowledge can assist in making more appropriate business decisions and gaining a better understanding of market behavior and trends, has been emphasized in the development of the concept of business intelligence and analysis.

By using machine learning techniques to stock up on medicines that have been found to be in high demand, a system is created in which there is no shortage of pharmaceuticals and medicines can be offered to individuals who need them . Health recommender systems can assist patients and healthcare practitioners in making better health-related decisions. Key drug shortages are anticipated to continue to be an issue. The pharmaceutical suggestion system, which is our goal, will be beneficial to the healthcare industry. People will not have to worry about running out of medicine since stores will be stocked far ahead of time because they will know which drugs are most likely to be purchased.

***Keywords*:** Machine learning, Apriori, Colab, recommendation

1 Introduction

COVID-19 carries an increased risk of serious consequences in some susceptible groups, such as the elderly, fragile, or those with several chronic illnesses. We can use such a categorization to put in place a method to combat medicine shortages. We strive to avoid shortages by using machine learning techniques to stock up on medicines that have been identified to be in high demand. Recently machine learning has been evolved from as a computational learning theory in artificial intelligence. It rose from an environment that was the integration of the interaction between available data, computing power, and statistical methodologies. Exponential growth of the available data compelled a spurt in computing power, which in turn stimulated the development of statistical methods to analyze large datasets.

Healthcare big data is a collection of patient, hospital, doctor, and medical treatment records that is so huge, complicated, scattered, and expanding at such a rapid rate that it is impossible to keep track of and analyze using typical data analytics methods[1]. To overcome these challenges, a big data analytics framework is used to apply machine learning algorithms to such a large quantity of data [2][3]. Technology has also progressed significantly in the discovery and development of novel pharmaceuticals that have the potential to benefit patients with complex illnesses. Given various amounts of accessible information about patients, a matching procedure for numerous different use cases has been developed[4]. Some large tech companies, such as IBM and Google, have developed machine learning tools to help patients find new therapy options. Precision medicine is an important concept in this discussion since it entails understanding mechanisms underlying complex disorders and developing new treatment options.

Although numerous semi-supervised strategies to give additional training data have been presented, automatically produced labels are typically too noisy to properly retrain models. As chronic diseases are long-lasting, it takes a lot of time to detect them[5]. As all of us know, medicine shortages were a major contributor to the enormous number of fatalities we saw in the pandemic. To combat this problem, our study suggests a solution by using machine learning techniques to stockpile medications that have been identified as being in high demand. This ensures that there is no shortage and we can provide them to people in need. Moreover, it helps to counter the problem of black-marketing of medicines since the most frequent ones are already present in sufficient stock, eliminating the need to buy them from dealers at exorbitant rates.

Big Data and the Cloud are two examples of new technologies that are helping to solve healthcare issues. Healthcare data is expanding at an exponential rate these days, necessitating an efficient, effective, and timely solution to cut mortality rates.

The importance of data collection, integration, processing, and reporting of underlying knowledge has been emphasized in the development of the concept of business intelligence and analysis, as well as how this knowledge can assist in making more appropriate business decisions and gaining a better understanding of market behaviors and trends. We have been able to unearth hidden information from data thanks to the massive expansion of data. Using current machine learning algorithms with minimal modifications, we may employ Big Data analysis for effective decision making in the healthcare industry. According to our findings, many academics are motivated to study machine learning algorithms in the health-care industry. However, selecting the appropriate algorithm to predict disease based on the data set generated by the researcher is always tough.

2 Proposed Model

The goal of this research is to use machine learning to help with drug supply. Using the Apriori algorithm's support metrics, the goal is to create a recommendation system for the medicine that a specific customer is most likely to buy, resulting in a win-win situation for both the customer and the shop owner: the customer gets the most appropriate medicine they want at all times and does not have to deal with the hassles of out-of-stock medicines; and the pharmacist learns the specific combination of medicines that is made available quickly. A lack of drug supply implies the medical black market is gone, which helps the economy thrive. The complete workflow of the proposed model is given in Fig.1.

Figure 1: Workflow of the model

Data Preprocessing
To support the laws and syntax that the specific ML model requires, the dataset must be preprocessed.
The following are the stages of preprocessing:

- Importing the desired libraries
- Importing datasets
- Dealing with missing data
- Encoding categorical data and encoding the dependent variable
- Feature scaling
- Splitting the dataset (training and test sets)

Apriori algorithm

- The apriori algorithm is an influential algorithm in determining frequent item sets for Boolean association rules.
- Apriori uses a “bottom up” approach, where frequent item sets are extended one item at a time(a step known as candidate generation, and groups of candidates are tested against the data.)
- Apriori is designed to operate on datasets containing transactions. E.g. – collection of items bought by customers, details of a website frequentation.

Working of Apriori model

The stages of the Apriori algorithm are given as follows and as depicted in Fig. 2.

1. Determine the itemsets' support in the transactional database and choose the lowest level of confidence and support.

2. Gather all of the dataset's support values that are greater than the minimum/selected support value.

3. Make a list of all the rules for subsets with a greater confidence value than the threshold or minimum confidence value.

4. Arrange the rules in order of decreasing lift.

5. The declining sequence of the lift will help us to better understand the relationship between the drugs.

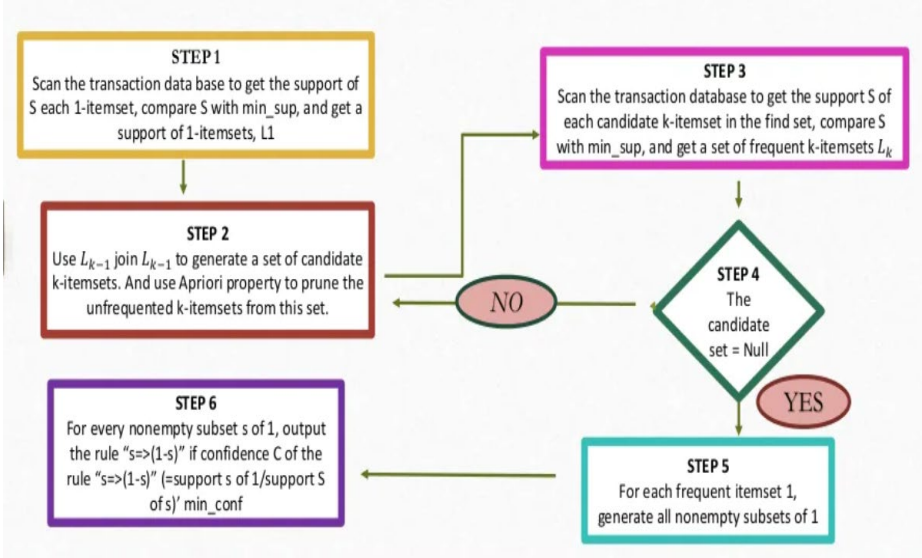

Figure 2: Working of Apriori model

Association rule learning

Association rule learning is a form of unsupervised machine learning approach that examines the reliance of one data item on another and maps appropriately to make it more lucrative. It tries to uncover some interesting relationships or links between the dataset's variables. It uses a set of rules to find interesting relationships between variables in a database.

The discovery of frequent itemsets in a transactions database is a crucial aspect of association mining. It's used in a lot of data mining activities that aim to uncover interesting patterns in datasets, such association rules, episodes, classifiers, clustering, and correlation, and so on.

Model description

In this project, we used the Apriori model to recommend the medicine combination that the customer is most likely to purchase. In 1994, R.Agrawal and Srikant introduced the Apriori technique[2], which uses recurring item sets to build association rules. It's designed to be used with transactional databases. These concepts can be used to determine how strongly or weakly particular items are connected.
The Apriori method uses a Hash tree and a breadth-first search to locate frequent items from a large dataset in an iterative fashion.

Association learning works on the if-then concept. The "If" element of association is called the Antecedent. The "Then" statement is called the Consequent. This type of relationship is called Single Cardinality. The metrics to find the association is given by **Support, Confidence and Lift.**
Support is referred to as the frequency of X, or the number of times an item appears in a collection. It is the proportion of the transaction T that contains the itemset X as defined in (1).

$$Supp(X) = \frac{Freq(X)}{T} \ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

Confidence can be defined as the frequency with which a rule is correct which is reflected in its degree of confidence. It's the ratio of a transaction that contains X and Y to the number of records that include X and Y as defined in (2).

$$Confidence = \frac{Freq(X,Y)}{Freq(X)} \ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

Lift is the ratio of the observed support measure and expected support if X and Y are independent of each other as defined in (3).

$$Lift = \frac{Supp(X,Y)}{Supp(X) \times Supp(Y)} \ldots\ldots\ldots\ldots\ldots\ldots \text{(3)}$$

It can have 3 values.

- ❖ Lift = 1: Antecedent and subsequent occurrence probabilities are independent of one another.
- ❖ Lift > 1: Determines the degree to which the two items are interdependent.
- ❖ Lift > 1: It indicates that one object is a replacement for another, implying that one

item causes harm to another.

Higher the lift, more is the association between those elements.

3 Simulations and Result analysis

The dataset used for simulation is a sample of medicine combinations that have been commonly bought by customers over the past 2 months. It is a random dataset that we have made to illustrate the idea of medicine prediction and contains 7500 example records.

The dataset has been randomly generated thus ensuring the accuracy of the model in the context of its probability of getting lucky for a particular dataset. Since it is generated randomly, it verifies the model's correctness in terms of its likelihood of being fortunate for a certain dataset. The practical use case of this dataset is that it will be given by the chemist shop based on their previous sales. The apriori algorithm will be executed on this for getting the preferred result.

The most commonly bought medicine items are shown in Fig.3

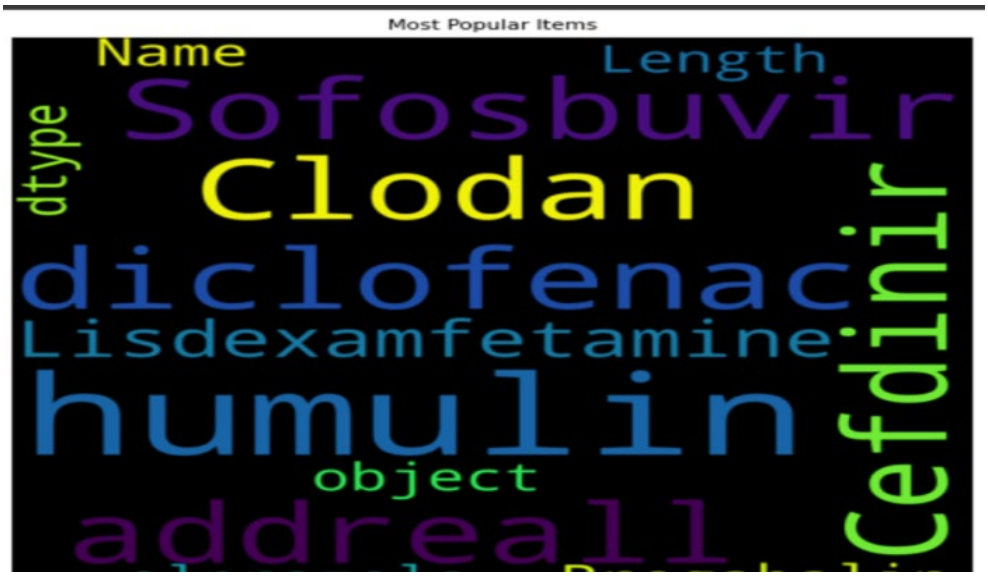

Figure 3: Word Cloud showing most popular items

Fig. 4 displays the most popular medicines as a frequency distribution. Fig. 5 is a representation of the results obtained by using the algorithm to predict most common associations, presented as a descending order of their Lifts. Table 1 shows the labels for the different medicine combination. Fig. 6 illustrates the association obtained for various medicine combinations, as recommended by the algorithm.

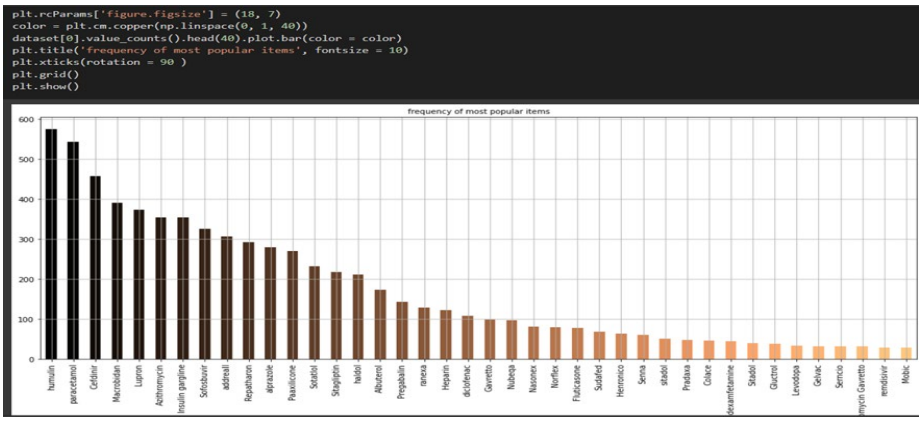

Figure 4: Frequency distribution of most popular items

Displaying the results ordered by descending lifts

```
[12] x=resultsinDataFrame.nlargest(n = 10, columns = 'Lift')
     x
```

	Left Hand Side	Right Hand Side	Support	Confidence	Lift
68	Levothyroxine	Lisdexamfetamine	0.004533	0.290598	4.843951
99	Rosuvastatin	Pregabalin	0.005733	0.300699	3.790833
105	Sotatlol	Sitagliptin	0.015998	0.323450	3.291994
103	Shringix	humulin	0.005199	0.254902	2.923577
77	sitadol	Lupron	0.005466	0.275168	2.886760
57	gabapentin	Insulin gargline	0.003733	0.482759	2.772720
29	Fluticasone	diclofenac	0.011332	0.224274	2.545056
101	Senna	Sitagliptin	0.006532	0.246231	2.506079
75	haldol	Lupron	0.016131	0.235867	2.474464
74	Sofosbuvir	Lupron	0.016664	0.233209	2.446574

Figure 5**:** Different medicine combination with their Support, Confidence and Lift Value

Table 1: Label for the medicine combination

Sl no.	**Left hand side**	**Right hand side**	**Label**
1	Evothyroxin	Lisdexamfetamine	Le+Li
2	Rosuvastatin	Pregabalin	Ro+Pr
3	Sotatlol	Sitagliptine	So+Si
4	Shringix	Humulin	So+hu
5	Sitadol	Lupron	Si+Lu
6	Gabapentin	Insulin garglin	Ga+In
7	Flucticasone	Diclofenac	Flu+dic
8	Senna	Sitagliptin	Se+Si
9	Haldol	Lupron	Ha+Lu
10	Sofosbuvir	Lupron	So+Lu

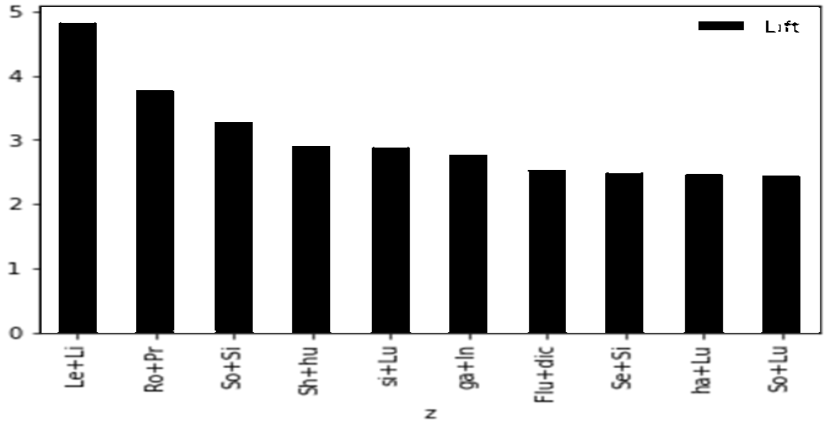

Figure 6:Variation of the Lift for the different medicine combination

Features of the Apriori algorithm

- Uses large itemset property
- Easily parallelized
- Easy to implement

Disadvantages of Apriori algorithm

- Assumes transaction database is memory resident
- Requires many database scans

4 Conclusions

Patients and healthcare providers can use health recommender systems to help them make better health-related decisions. Shortages of key medicines will likely continue to be a problem. Our objective, the medicine recommendation system will be helpful for the healthcare sector. The fundamental goal of this strategy is to ensure that the general public has access to the greatest possible variation of pharmaceuticals accessible on the market at all times. Based on prior sales of those drugs, our model will only offer the best related combination of medicines that go together in a certain way. As a result, the patient and the chemist store can have the best possible transaction. On the other hand, it enables the chemist to update his or her stock to its maximum potential at any time, ensuring that the patient receives the finest possible versions of the drugs whenever he or she requires them. In future this apriori based machine learning recommendation model can be expanded to allow low infrastructural casualties in a healthcare center as it will always ensure that the best possible medicine or other health equipment are available at all times of the year . This will boost the lack of technical and managerial policies that are lacking today in different healthcare centres across India. This model can be further integrated with UI/UX apps which will allow a patient and his/her family to get a clear visual understanding of the current status of the different healthcare facilities that are available at a healthcare center in some developed areas without even travelling long distances in search of a preferable diagnostic centre for the pa-

tient . This approach is expected to save many lives and thereby contribute to a better policy making attitude for the common people.

References

1. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst*. 2014;2:3. Published 2014 Feb 7. doi:10.1186/2047-2501-2-3
2. Al-Maolegi, Mohammed & Arkok, Bassam. (2014). An Improved Apriori Algorithm For Association Rules. International Journal on Natural Language Computing. 3. 10.5121/ijnlc.2014.3103.
3. Tran, T.N.T., Felfernig, A., Trattner, C. *et al.* Recommender systems in the healthcare domain: state-of-the-art and research issues. *J IntellInfSyst* **57,** 171–201 (2021).
4. Han, Q., Ji, M., Martínez de Rituerto de Troya, I., Gaur, M., &Zejnilovic, L. (2018). A hybrid recommender system for patient-doctor matchmaking in primary care.In *The 5th IEEE international conference on data science and advanced analytics (DSAA), (pp. 1–10)*.
5. Pahulpreet Singh Kohli and Shriya Arora. Application of machine learning in disease prediction. In 2018 4th International Conference on Computing Communication and Automation (ICCCA), pages 1–4. IEEE, 2018.
6. Munira Ferdous, Jui Debnath and Narayan Ranjan Chakraborty, Machine Learning Algorithms in Healthcare: A Literature Survey, In, that on 2020 11th International Conference on Computing, Communication, and Networking Technologies (ICCCNT)
7. Shweta Ganiger and KMM Rajashekharaiah. Chronic diseases diagnosis using machine learning. In 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), pages 1–6. IEEE, 2018.
8. Dharavath Ramesh, PranshuSuraj, and LokendraSaini. Big data analytics in healthcare: A survey approach. In 2016 International Conference on Microelectronics, Computing and Communications (MicroCom), pages 1–6. IEEE, 2016
9. Geron, “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow”, O’Reilly Media, Inc., Canada, 2019
10. D. Ravì et al., "Deep Learning for Health Informatics,", IEEE Journal of Biomedical and Health Informatics, 21(1), 2017, pp. 4-21.
11. Jupyter Notebook: Project Jupyter. Accessed November 2021.[Online]. Available: http://jupyter.org
12. Tran, T.N.T., Atas, M., Felfernig, A., Le, V.M., Samer, R., Stettinger, M. (2019).Towards social choicebased explanations in group recommender systems. In Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, UMAP ’19, (pp. 13–21). Association for Computing Machinery, New York, NY, USA.
13. Dave DeCaprio, Joseph Gartner, Carol J. McCall, Thadeus Burgess, KristianGarcia, Sarthak Kothari, Shaayaan Sayed, Building a COVID-19 vulnerability index, Journal of Medical Artificial Intelligence,2020
14. Ahuja, Vanita& Nair, LekshmiV. (2021). Artificial Intelligence and technology in COVID Era: A narrative review. Journal of Anaesthesiology Clinical Pharmacology. 37. 28. 10.4103/joacp.JOACP_558_20.