Graduate Research in Engineering and Technology (GRET)

Volume 1
Issue 6 *Application of Intelligent Computing and Big data Analytics in Healthcare*

Article 2

May 2022

# A Machine Learning based Drug Recommendation System for Health Care

Mahima Mohapatra
*Institute of Technical Education and Research, Bhubaneswar, Odisha, India*, mahimamohapatra1@gmail.com

Mamata Nayak
*Institute of Technical Education and Research, Bhubaneswar, Odisha, India*, mamatanayak@soa.ac.in

Saswati Mahapatra
*Institute of Technical Education and Research, Bhubaneswar, Odisha, India*, saswatimohapatra@soa.ac.in



## Recommended Citation

Mohapatra, Mahima; Nayak, Mamata; and Mahapatra, Saswati (2022) "A Machine Learning based Drug Recommendation System for Health Care," *Graduate Research in Engineering and Technology (GRET)*: Vol. 1: Iss. 6, Article 2.
DOI: 10.47893/GRET.2022.1109
Available at: https://www.interscience.in/gret/vol1/iss6/2

# A Machine Learning based Drug Recommendation System for Health Care

[1]Mahima Mohapatra, [2]Mamata Nayak, [3]Saswati Mahapatra

[1,2,3] Institute of Technical Education and Research, Bhubaneswar, Odisha, India

Email: mahimamohapatra1@gmail.com,mamatanayak@soa.ac.in, saswatimohapatra@soa.ac.in

**ABSTRACT**

In today's digital era healthcare is one among the major core areas of the medical domain. People trying to find suitable health-related information that they are concerned with. The Internet could be a great resource for this kind of data, however you need to take care to avoid getting harmful information. Nowadays, a colossal quantity of clinical information dispersed totally across different websites on the Internet prevents users from finding useful information for their well-being improvement. Errors in medication are one of the foremost severe medical faults that would be a threat to patients' lives. These problems increases the requirement to use recommendation systems within the domain of healthcare to assist users create additional economical and correct health-related decisions. During this paper, drug recommendation systems are developed to help end-users in distinctive correct medications for a particular wellness based on the reviews of other end-users provided on totally different medications for various specific diseases. The goal of this recommendation system is to examine the dataset using data mining concepts, visualization, sentiment analysis and recommend drugs based on the condition, ratings and reviews using Machine Learning approaches, Content and Collaborative filtering approach, for each health condition of a patient.

**Keywords:** Drug, Recommendation System, Content Filtering, Condition, Machine Learning.

## I. Introduction:

One of the most widely intensive topics on the internet is health-related information. Considering the present state of affairs everywhere the globe, people are more and more concerned about health and medical diagnosis issues. Some of the survey research like the Pew Internet survey says that 55% of the people having Internet, have used the Web to get health-related information [1]. There is a study that focused to direct the actual commonness of searches related to health on the internet by analyzing search terms that people entered into common search engines. Also, to make some preliminary efforts in approximately describing and classifying these searches [2]. According to the write-up found in NCBI, it is reported that annually around 99,000 people die, because of mistakes done by medical professionals in hospitals in a ratio of 1:5 doses [3]. These problems raise the requirement to use recommendation systems in the field of healthcare to assist end-users make more efficient and accurate health-related decisions. A recommendation system is very essential in this fast-growing technological world, which can save lives. In this paper, the proposed drug recommendation system and its working are depicted. In the system, the drug is offered on a specific condition dependent on patient reviews and ratings using technologies like Machine Learning, Data mining, etc. It is based on Content and a Collaborative filtering approach. The basic objective of the Drug Recommendation system is to design an effective and accurate system for recommending drugs for patients. As an outsized range of datasets are available over internet sources, our system will analyze the data and aims to fulfil the objective with accuracy, scalability, and efficiency.

## II. LITERATURE REVIEW

Recommender systems aim to predict users' interests and recommend product items that are quite seemingly fascinating for them. Numerous recommendation framework techniques have been anticipated since the mid-1990s, and many types of recommendation system software packages are developed recently for a variety of applications in different industries. With this framework, we are able to determine industries that are capable of acquiring advantages from recommendation systems. E-Commerce Business where recommendation systems were initially widely used. Media businesses are one among the first to get into recommendations. Without a

recommendation system, its tough to examine news sites. Banking, Small and mid-size enterprises are main for recommendations. Knowing a end user's detailed monetary affairs , along with their past choices and preferences, in inclusion to information of thousands of similar users is quite powerful. Let's discuss the connected papers and analysis of drug recommendation systems. Table 1 represents the papers for the initial and final section similarly because of the rate of included papers in percent for different resources. Also, a column per phase for the outcome of the searches including Google Scholar searches to make further transparency. The drug recommendation system Galen OWL [5] used the Greek drug for prediction and finds details on the medication and extra data, like interactions with alternative medication. The data mining and machine learning-based approach by Sun et al. [6] analyzed electronic medical data to determine typical cure regimens and measure the efficacy of these regimens for specific patient groups. There could be another drug recommendation system that can be cloud-based platform utilizing numerous algorithms [7]. There are more research papers accessible for a distinct purpose within the Healthcare domain [4].

## III. PROPOSED METHODOLOGY

We worked on Drug Review Dataset and created the drug recommender model using Google Colab platform. First, the dataset is put through pandas, numpy, etc libraries for cleaning, preparation and visualization. Then, we tend to use it for the process that uses techniques like Stemming using NLTK Library, the count vectorizer method. We tend to check the accuracy of the dataset using logistic regression, evaluating models with numerous Similarities and Distance Matrices. Finally , Building a content filtering-based and a collaborative filtering-based recommendation system. The framework of the workflow of the proposed drug recommender model is shown in Fig. 1.

TABLE I. NUMBER OF PAPERS FOUND IN THE RESPECTIVE SEARCH ENGINES (INCL. GOOGLE SCHOLAR RESULTS AND FINAL INCLUSION RATE)

| Resource | Initial Total (incl. Google Scholar) | Initial Google Scholar only | Final Total (incl. Google Scholar) | Final Google Scholar only | Inclusion rate in % Total (incl. Google Scholar) |
|---|---|---|---|---|---|
| ACM | 22 | 2 | 3 | 0 | 14% |
| IEEExplore | 7 | 5 | 5 | 4 | 71% |
| ScienceDirect | 5 | 0 | 0 | 0 | 0% |
| Elsevier | 5 | 0 | 0 | 0 | 0% |
| John Wiley Inc. | 3 | 0 | 0 | 0 | 0% |
| Springer | 1 | 1 | 1 | 1 | 100% |
| Americana Medical Informatics Association | 1 | 1 | 1 | 1 | 100% |
| Journal of Biomedical Semantics | 2 | 2 | 2 | 2 | 100% |
| International Journal of Environmental Research and Public Health | 1 | 1 | 0 | 0 | 0% |
| Journal of Chemical and Pharmaceutical Sciences | 1 | 1 | 0 | 0 | 0% |
| Advanced Internet of Things | 1 | 1 | 1 | 1 | 100% |
| Clinical Pharmacology & Therapeutics | 1 | 1 | 0 | 0 | 0% |
| International Journal of Computer Applications | 1 | 1 | 0 | 0 | 0% |
| Total | 52 | 17 | 13 | 9 | 25% |

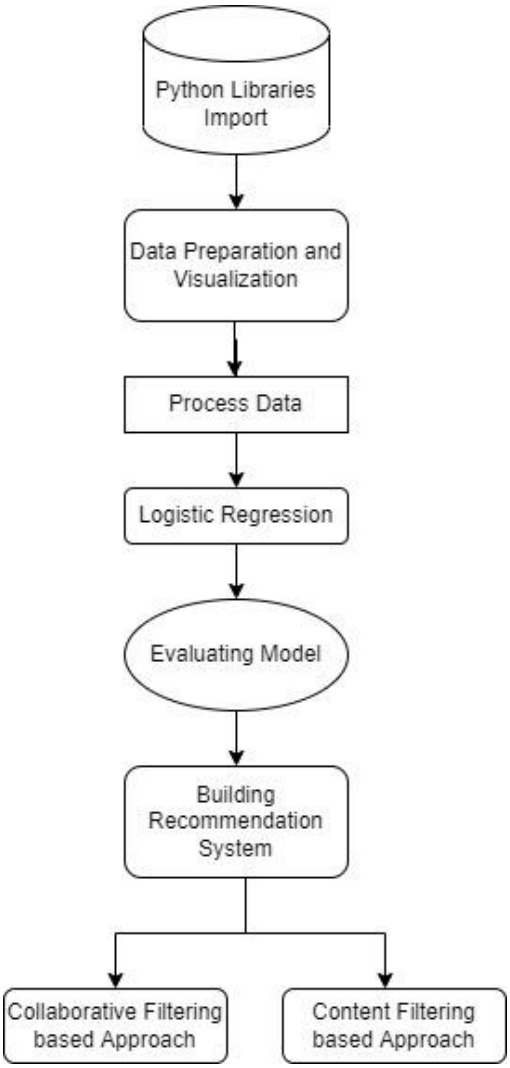


**Fig 1: Proposed workflow**

The main steps in providing the recommendation for the condition are as follows:

**Step 1: Importing Necessary Python Libraries**

In this step, the necessary python libraries are imported. Pandas is a data analysis tool that allows the extraction of data. Numpy facilitates advanced mathematical operation on large numbers of data. Pyplot is a module in matplotlib that manipulates elements of a figure. Word Stemming libraries using NLTK libraries like Word Net Lemmatizer and Porter Stemmer are used to make root words in a string. Count Vectorizer is used to make words in vector form. Logistic regression is used for the understanding of the accuracy of the dataset and the validity of the recommendation. Other Similarity and distance metrics Libraries like person, spearman, Jaccard score, etc are used for evaluating the model.

**Step 2: Data Extraction and Exploratory Analysis**

We used Drug Review Dataset in two files (test and train) directly downloaded from Blob Storage in Azure Cloud. The Dataset contains a unique ID, drug name, condition, review, rating, date and user count. Then, we convert the 'Rating' and 'Useful Count' into a Numeric data type for generating a new column of 'most reviewed' drugs. The top 10 Most Reviewed Drugs is shown in Fig 2 where we can analyze the number of reviews for the 10 drugs. The graphical representation for the total number of drugs per condition (Top 20) and the number of reviews in each year is illustrated in Fig 3 and4 respectively. Figure 5 demonstrates the d by highest monthly reviews for different drugs.

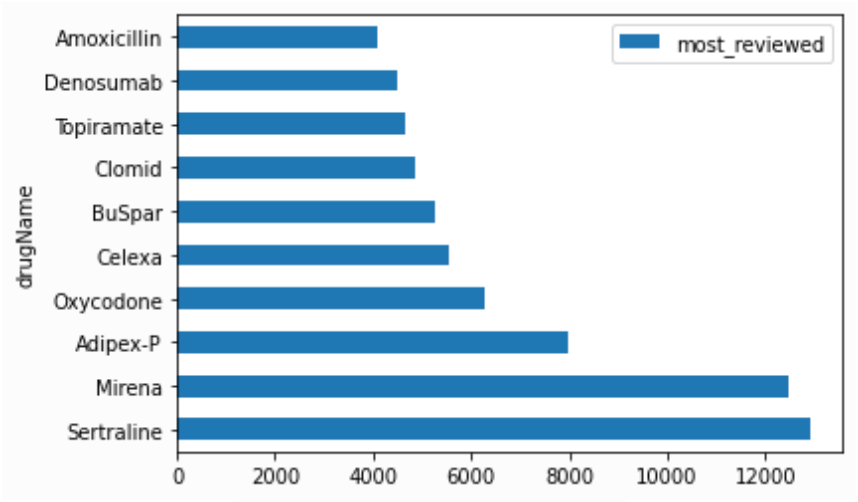


Fig 2: Top 10 Most Reviewed Drugs

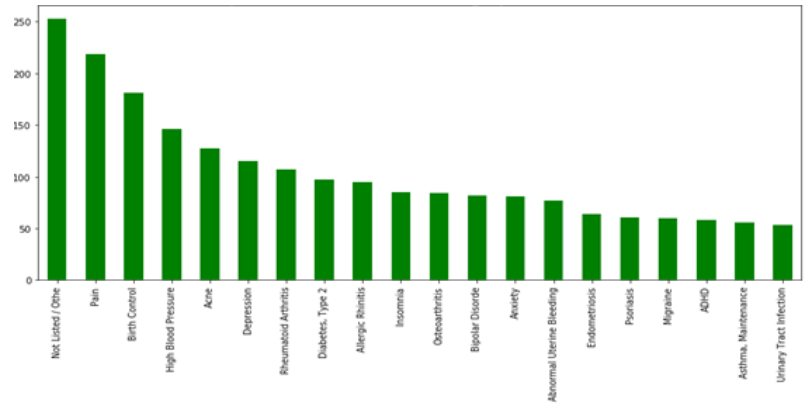


Fig 3: Top 20 number of drugs per condition

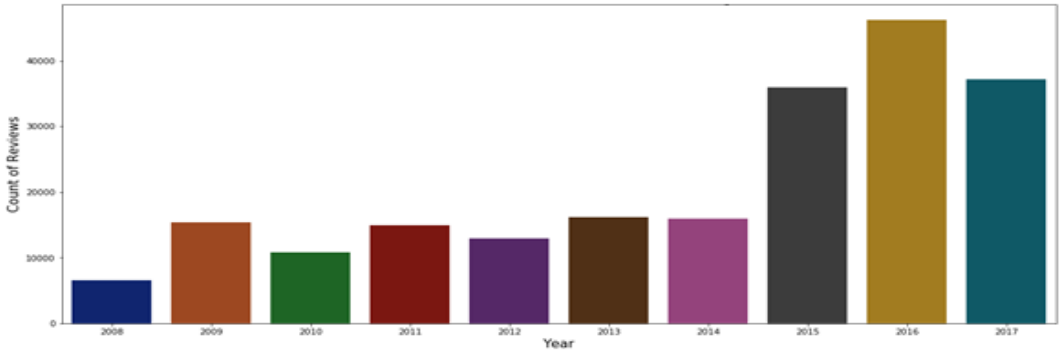


Fig 4: The Number of Reviews in each year

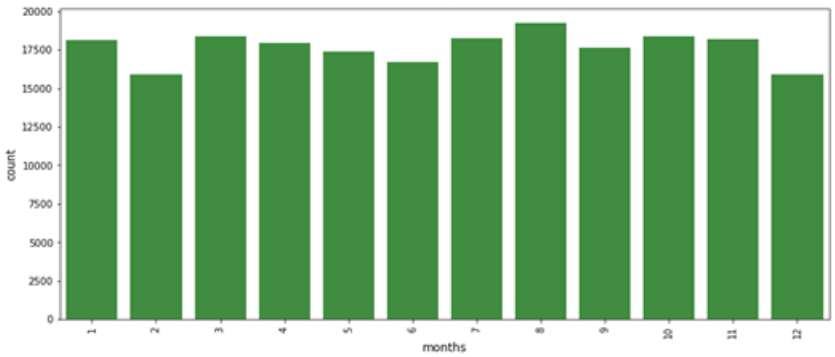


Fig 5: Month wise the number of reviews

We analyze and visualize the data in the Drug Review dataset in various factors and graphs for easy and qualitative understanding of the data for further process. Next, in the 'Review' column, it contains the reviews of people on various drugs of various conditions in string format. We remove the unwanted characters and split in words from the string values. We then concatenate Drugs and Condition values with Reviews to build a combined metadata of drugs.

**Step 3: Process Data**

The Drug review dataset is splitted into in two files for training and testing respectively. The whole data is processed to remove irrelevant information.

**1)Stemming using NLTK Library:** Before going to this step, a new data frame was created with attributes like unique ID, drug name, condition and tags. Column 'tags' consists of concatenated values of a drug name, condition and reviews. We perform Stemming on the new data frame. Stemming using NLTK Library is a technique used to extract the base form of the words by removing affixes from them. Porter Stemmer is applied to check root words. Word Net Lemmatizer is also deployed to check root words and meaning consistency.

**2)Implementing Count Vectorizer to make words in vector form:** In this step, column 'tags' is considered. Count vectorizer is used to make the words in vector form. Maximum features is set to 500 for removing a d stop words from a string. Applied Fit transform to Lemma processed vector as it gives more relevant meaning.

**Step 4: Logistic Regression**

In this step, to understand whether the dataset is a useful dataset and powerful, logistic regression is applied to identify the new recommendation from this dataset is valid or not. Basically, it checks the accuracy of the dataset. First, we import the Logistic Regression module and create a classifier object using the function Logistic Regression(). Then, using fit() function, we fit the model on train dataset. After that using predict() function, we perform prediction on test dataset. Then, we use confusion matrix for model evaluation. A confusion matrix is a table that is used for the evaluation of performance of a classification model. The model was evaluated using model evaluation metrics such as accuracy, precision, F1 score, and recall. Compiled all of the metric into a data frame in Fig 6. The figure shows that it has a classification rate of 80% which is considered as good accuracy.

| | LogReg1 |
|---|---|
| Accuracy | 0.805 |
| Precision | 0.823 |
| Recall | 0.914 |
| F1 Score | 0.866 |

Fig 6: Performance of Logistic regression on Drug review data set.

**Step 5: Evaluating Model with Various Similarity and Distance Metrics**

In this step, we evaluated the model with various Similarity and Distance Metrics like Pearson Correlation, Spearman Correlation, Kendall Tau's Correlation, Cosine Similarity, Jaccard's Similarity, Euclidean distance and Manhattan Distance with first drug and random drugs.

**Step 6: Building Recommendation System**

There are various approaches to build a recommendation system. We opted with two approaches for:

**Content-based filtering:** A content-based recommender works with information that is provided by the user, either explicitly (rating) or implicitly (clicking on a link). In this method, we by choosing cosine similarity for recommender system building. Then, sorting similarities in descending orders to make more similar drugs on top. Defined a function to recommend a drug based on the condition selected.

**Collaborative-based filtering:** In collaborative filtering, it finds similar users and recommend what similar users like. It recommends the items based on the reviews of the previous users. In this method, we are comparing for similar condition by building User-item interactions Matrix and compare for similar drug. User-item matrix is a basic foundation of traditional collaborative filtering techniques. We sorted the values according to the rating. Testing the drug matrix which will indicate the medicine which is suitable for the input condition and analysed correlation with similar condition.

**IV. RESULTS**

This section contains the experimental result and findings of a machine learning-based drug recommendation system. Google collab platform with its built-in python libraries was used for creating the drug recommender model. Various machine learning techniques like Stemming, Logistic regression, etc are used for the extraction of data and finding its accuracy and usefulness. We

built a content filtering-based system that takes input conditions and recommends drug for the specific input condition. Then, a collaborative filtering-based system that takes input condition and indicates the drug for that input condition and also gives analysis correlation with similar condition.

## V. CONCLUSION

In this paper, a machine learning based drug recommendation system using content based and collaborative based filtering approach was developed using a sample dataset from Kaggle over Google Colab platform. In this current situation, this recommendation system can be helpful and effective for both healthcare professionals and patients. Furthermore, we can develop this system with more advanced machine learning algorithms to be able to take care of the exact needs of a patient to save as many lives as possible. Advancement in technology can really have a great impact over healthcare industry.

## REFERENCES


1. Pew Internet and American Life Project. The Online Health Care Revolution: How the Web helps Americans take better care of themselves. 11-26-2000.

2. G. Eysenbach and Ch. Kohler. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the Internet.

3. Iyer MH. Medical errors in top 10 killers: WHO, 2011. [Last accessed on 2012 May 18]. Available from: http://articles.timesofindia.indiatimes.com/2011-04-20/india/29450526_1_medicalerrors-patient-safety-negligence .

4. A Literature Review on Medicine Recommender Systems Benjamin Stark, Constanze Knahl, Mert Aydin, Karim Elish. Department of Computer Science, Florida Polytechnic University, Lakeland, USA

5. Doulaverakis, C., Nikolaidis, G., Kleontas, A., and Kompatsiaris, I. 2012. GalenOWL: Ontology-based drug recommendations discovery. Journal of Biomedical Semantics 3, 14

6. Sun, L., Liu, C., Guo, C., Xiong, H., and Xie, Y. 2016. Data-driven Automatic Treatment Regimen Development and Recommendation. KDD '16 Proceedings of the 22nd ACM SIGKDD International. Conference on Knowledge Discovery and Data Mining, 1865–1874.

7. N. Komal Kumar and D. Vigneswari. A Drug Recommendation System for Multi-disease in Health Care using Machine Learning.

8. Zihayat M, Ayanso A, Zhao X, Davoudi H, An A(2019) A utility-based news recommendation system, Decis Support Syst 117:14-27