International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET231036

Recommendation of Drugs and Its Substitute Medicines Also Purchasing Using Cosine Similarity Vector

K. Saranya[1], R. Jayashree[2], A. Ranjitha[2]

[1]Assistant Professor, Information Technology, Kings Engineering College, Tamil Nadu, India

[2]Information Technology, Kings Engineering College, Tamil Nadu, India

ARTICLEINFO

Article History:

Accepted: 20 April 2023
Published: 07 May 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
28-33

ABSTRACT

In the current digital era, healthcare is one of the main focuses of the medical sector. A mistake with a patient's medication is one of the most dangerous medical errors that might risk the patient's life. It leads the healthcare sector to assist users in making more suitable and cost- effective health-related decisions. A machine learning-based drug recommendation system that takes into account the patient's reported symptoms or drugs is suggested by this study. The algorithm utilised is cosine similarity vector and data frames of medicine, and the system uses supervised learning techniques such as decision trees and K- nearest neighbours to propose the alternative drug and urges users to buy that specific drug.

Keywords: Drug, Recommendation System, Decision Tree, K-Nearest Neighbour, Cosine Similarity Vector, Machine Learning.

I. INTRODUCTION

One of the most searched-for subjects on the internet is health-related information. Considering the current global circumstances, Healthand worries about medical diagnoses are causing people more worry on a regular basis. 55% of Internet users have used the Web to access health- related information, according to certain survey studies, including the Pew Internet study[1].One study examined the search terms that people entered into well-known search engines to ascertain the actual frequency of searches for health-related subjects on the internet. Additionally, to make some first attempts at broadly synthesizing and categorising these questions[2]. According to the article published in NCBI [3], around 99,000 people every year perish away as a result of mistakes committed by hospital staff members in a ratio of 1:5. These problems underscore the need for recommendation systems in the healthcare sector to assist end users in making more accurate and effective decisions regarding their health. A recommendation system is essential in today's rapidly changing technological environment and may possibly save lives. This document describes the proposed medicine recommendation system and how it works. This project will suggest medications and offer outcomes depending on user requirements. Additionally, it can provide prescriptions for cheaper drugs while saving time and money by automating the drug suggestion process[12]. It can also enhance patient outcomes by providing individualised

medicine recommendations that are connected to an online consumer healthcare system. Additionally, users or patients can purchase the medication on that site. that are customised for the patient's specific symptoms and medical history. This approach has the potential to significantly improve medical care quality and provide patients with more individualised treatment alternatives.

II. LITERATURE SURVEY

The goal of recommender systems is to anticipate users' interests and make suggestions for products that seem particularly exciting. Since the middle of the 1990s, many distinct recommendation framework strategies have been expected, and recently, numerous types of recommendation system software packages have been produced for several applications in various industries. With the use of this framework, we can identify the sectors that stand to benefit from recommendation systems[5]. Recommendation systems were first widely employed in the e-commerce industry. Among the first to enter the recommendations market are media companies. Examining news sites is challenging without a recommendation system. Banking and small to medium- sized businesses are the main sources of advice. Knowing a user's complete financial situation, past choices and preferences, and the data of thousands of other users who are like them are quite useful.Two deep fusion models for assessing medication evaluations based on 3-way decision theory are suggested by Mohammad Ehsan Basiri at el[4]. The first fusion model, 3W1DT, was created as a primary classifier using deep learning and a traditional learning process as a backup method used when there is insufficient confidence in the deep method when labelling test samples. 3W1DT is a 3-way fusion of a single deep model with a traditional training algorithm. The second recommended deep fusion model learns the 3-way fusion of three deep models with a traditional model (3W3DT), three deep, and a traditional model throughout the whole training set, and each classifies the test sample uniquely. The most reliable classifier is then selected to categorize the test drug review. A drug recommendation assistant, constructed using machine learning is introduced by Vikrant Doma at el[6], methods as well,which generates its accuracy with several major datasets. The introduced method allows the contrasting effects, feedback, rankings, to be manifested and then recommends the most"effective" medication for a given Person.Machine learning algorithms are used such as linear SVC, ridge classifier, Naive bayes. This project recommends medicine in two languages. An effort has been made to include machine learning and deep learning techniques into recommender systems because of the significant increase in AI development. These days, recommender systems are widely used in the restaurant, e-commerce, and travel sectors, among other industries. regrettably, there aren't many studies in the field of drug proposal framework using sentiment analysis because medication reviews are much more difficult to analyse because they contain clinical terminology like names of infections, reactions, and synthetic names that are used to make the drug[7].A decision making process is ineffective in hospitals, so that the association rule mining algorithm employed to find medicine association rules. In addition to that graph mining and clustering algorithm, louvian algorithm is used. In this project the relationship between drugs are being extracted, extraction of graph and clusters will produce an enriched recommendation via ATC code. KNN is extremely easy to use and yet carries out complex classification tasks in its most basic form. Rather, all data is used for training when a new data point or instance is classified. KNN is an algorithm of non-parametric learning, which means the underlying data are unassuming[8]. Health recommender systems have emerged as tools to support patients and healthcare professionals to make better health-related decisions. In this paper, we have given insights into recommendation scenarios offered by these systems, such as food and drug recommendation, health status

prediction, physical activity recommendation, and health care professional recommendation. For each recommendation scenario, various algorithms have been employed, which are based on recommendation techniques[9].The SVM notion, one of the most trustworthy and efficient machine learning algorithms, is based on the Structural Risk Minimization principle of computational learning theory. According to this theory, data is assessed and hyperplanes are used to represent decision limits. It uses four kernel structures for classification tasks, including linear, polynomial, radial-based, and sigmoid functions, when the input data cannot be easily separated[10]. This is done by mapping the input data into a high- dimensional feature space. Each class's text vectors are divided by the hyperplane in a way that preserves as much of the distinction as feasible.On the basis of treatment data, Xiaohong Jiang et al[15]. investigated three different algorithms: the decision tree method, the support vector machine (SVM), and the backpropagation neural network. Model exactness, model proficiency, and model variety were the three distinct boundaries that were used to choose SVM for the medicine suggestion module.Additionally, a method for checking for errors was suggested to guarantee the accuracy of the administration, analysis, and data.

III. PROPOSED METHODOLOGY

Drug suggestions are made using the medication dataset. The name of the medication used (text), the patient's review (text), and the patient's condition (text) are three of the six variables in this dataset.

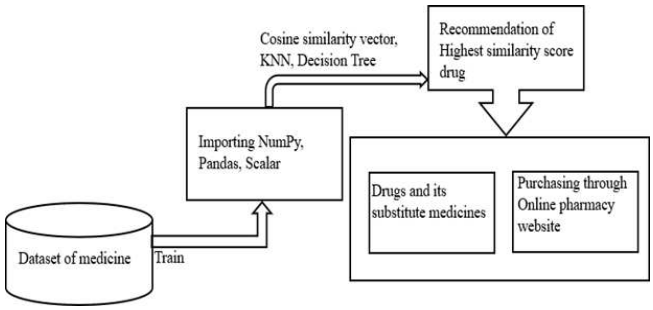

Fig.1 Workflow of drug recommendation

The dataset includes the drug and patient dataset, which suggests medicines based on the correlation between patient symptoms. Additionally, a pharmacy website where the user may purchase the drug will be accessible underneath the list of recommended medications.

A. Data cleaning

Data cleaning, often known as data cleaning or scrubbing, is the process of identifying and removing errors, duplicate data, and extraneous data from a raw dataset. Data purification, a step in the data preparation process, enables correct, tenable data to produce accurate visualizations. Standard data preparation approaches were used in this study, including checking for null values, deleting duplicate rows, removing superfluous values, and removing text from rows. Furthermore, the data in the dataset is cleansed and accurate medicine data is processed and collected.

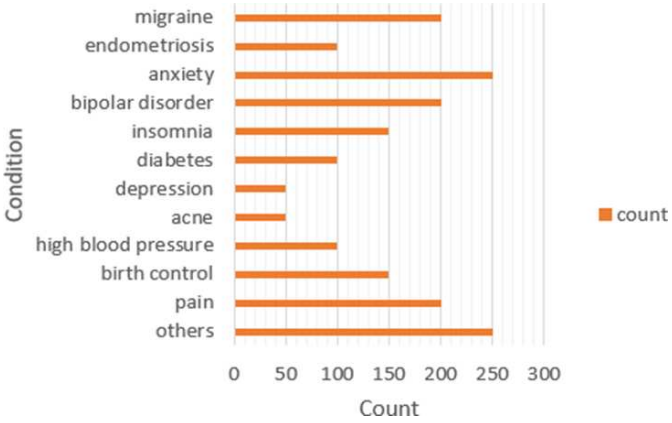

Fig.2 Medical conditions and available drugs

A series of preprocessing procedures required to both reduce the dataset so that it has the essential data points and attributes for the analysis and to enable the machine learning system and algorithms to read and analyse the data. The creation or measurement of extra qualities from the data may also be significant if they may improve analysis and lead to more accurate predictions.

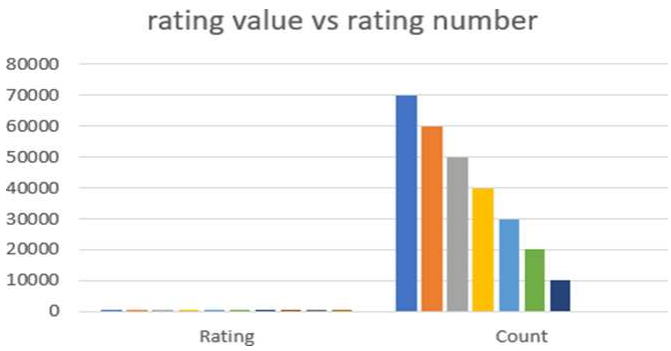

Fig.3 count of rating value vs rating number

B. Training the dataset

In this step, the libraries which is used in recommendation is trained and imported, those libraries are Pandas, Numpy, Scalar. An open-source library called Pandas is described as offering high-performance data manipulation in Python. Panel Data, which denotes econometrics from multidimensional data, is where the term Pandas originates. In Python, it is utilized for data analysis. With NumPy, array objects should be up to 50 times quicker than Python lists traditionally. The array object in NumPy is known as ndarray, and it has several auxiliary methods that make working with ndarray very simple. In data science, where efficiency and availability of resources are crucial, arrays are utilised often. The technique of normalising the variety of characteristics in a dataset is known as feature scaling. Real-world datasets frequently include characteristics that vary in size, scope, and units. We must thus do feature scaling in order for machine learning models to comprehend these characteristics on the same scale.

C. K-Nearest Neighbour

Even in its most basic version, KNN can do difficult classification jobs and is quite simple to use. The algorithm is inefficient because the learning phase lacks experience. Instead, every piece of data is used for training whenever a new instance or data point is identified. The underlying facts are unimpressive since KNN is a non- parametric learning technique [11]. Since most real-world data do not adhere to theoretical presumptions like linear separateness, uniform distribution, etc., this feature is quite helpful. This accurately estimates the variation between a single sample and each training sample. Then it selects the K nearest data samples and assigns the data sample to the category where most of the K data samples are present [13].

D. Cosine similarity score

A measurement of how similar two data points on a plane are is called cosine similarity. Cosine similarity is used as a metric in various machine learning algorithms, such as the KNN for calculating the distance between neighbors, in recommendation systems to suggest films with similar themes, and for textual data to identify textual similarity in a document. One such algorithm that operates based on content similarity is the recommendation system used in machine learning. There are many different methods for determining how similar two pieces of material are, and recommendation systems essentially utilise the similarity matrix to suggest comparable pieces of content to the user depending on the user's accessing preferences. Thus, any recommendation data can be obtained, and the necessary attributes that are helpful for suggesting the contents may be extracted from the data. The appropriate textual data must be made accessible before using the CountVectorizer to vectorize the text in order to produce the similarity matrix. The cosine similarity metrics of Scikit Learn may therefore be used to suggest users after the similarity matrix has been produced. For the selected textual data for recommendations, the cosine similarity would provide a similarity matrix, and the material with greater similarity scores may be sorted using lists. Here, cosine similarity would consider the often-occurring phrases in the textual data, the terms' greater frequency vectorization, and the content's higher recommendation percentages. So this is how recommendation algorithms use cosine similarity.

condition	drugname	score
Pain	Neurontin	0.1584667
Pain	Nortriptyline	0.1717711
High Blood Pressure	Adalat CC	0.3031911
High Blood Pressure	Zestril	0.3058551
Acne	Atralin	0.0885455
Acne	Retin A Micro	0.0973992
Birth control	Ortho Micronor	0.0061499
Depression	Remeron	0.1246011
Depression	Provigil	0.2401855

Fig.4 Similarity scorers of medicines with similar symptoms

E. Drug recommender system

The most accurately anticipated results were chosen and pooled to create the combined forecast after the metrics were evaluated. The combined findings were then multiplied by the normalized useful count to get a final score for each medicine for a given condition. The quality of the medicine improves with a higher score[14]. Looking at the distribution of helpful count and analyzing how much difference there is between the least and most extreme values—roughly 1300—led to the standardization of the user count. The variance, which is 36, is also very large. The rationale behind this is that more people will read the survey as more drugs are searched for, whether their reviews are favourable or negative, increasing the poll's usefulness. To normalize usable count by circumstances, we did so when developing the recommender system. In addition to that, the recommended drugs will have the substitutes as alternative drug for the searched drug, where the user can find it in helpful way. Furthermore, the pharmacy link will be provided to purchase the drug in an online pharmacy platform.

IV. CONCLUSION

In this project, the Jupyter platform and a sample Kaggle dataset were used to create a machine learning-based medication recommendation system. In addition, a cosine similarity score, a decision tree algorithm, and KNN were used to determine which medications had the highest similarity scores. The pharmaceuticals with the greatest similarity score will be suggested, and the user can purchase those pills from the online pharmacy provided below the proposed drug. To be able to accommodate each patient's unique needs, we may upgrade this system with more advanced machine learning techniques.

V. REFERENCES

[1]. Fox, S.(2000). The online health care revolution: How the web helps Americans Take better care of themselves. A Pew Internet and Americans Life Product Online Report.

[2]. Eysenbach, G., & Kohler, C. (2003). What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the internet. In AMIA annual symposium proceedings (Vol. 2003, p. 225). American Medical Informatics Association.

[3]. Iyer, M. (2011). Medical errors in top 10 Killers: WHO. The Times of India.

[4]. Basiri, M. E., Abdar, M., Cifci, M. A., Nemati, S., & Acharya, U. R. (2020). A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. Knowledge- Based Systems, 198, 105949.

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

[6]. Doma, V., Singh, S., Arora, N., Ortiz, G., Saran, P. K., Chavarin, S., & Pirouz, M. (2020). Automated drug suggestion using machine learning. In Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), Volume 2 (pp. 571- 589). Springer International Publishing.

[7]. T. N. Tekade and M. Emmanuel, "Probabilistic aspect mining approach for interpretation and evaluation of drug reviews," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, 2016, pp. 1471-1476, doi: 10.1109/SCOPES.2016.7955684

[8]. Gazalba, I., & Reza, N. G. I. (2017, November). Comparative analysis of k- nearest neighbor and modified k-nearest neighbor algorithm for data classification. In 2017 2nd international conferences on information technology, information systems and electrical engineering (ICITISEE) (pp. 294-298). IEEE.

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

[10]. Y. Chen, J. Xiong, W. Xu, and J. Zuo, "A novel online incremental and decremental learning algorithm based on variable support vector machine," Cluster Comput., vol. 22, no. 3, pp. 7435–7445, May 2019, doi: 10.1007/s10586- 018-1772-4.

[11]. Okfalisa, I. Gazalba, Mustakim, and N. G. I. Reza, "Comparative analysis of k- nearest neighbor and modified k-nearest neighbor algorithm for data classification," in Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017, Feb. 2018, vol. 2018-January, pp. 294–298, doi: 10.1109/ICITISEE.2017.8285514.

[12]. Chiang, W.H., & Ning, X.Drug recommendation toward safe polypharmacy (supplementary materials). bioRxiv, 2019.

[13]. J. Jiang, Y. Chen, X. Meng, L. Wang, and K. Li, "A novel density peaks clustering algorithm based on k nearest neighbors for improving assignment process," Phys. A Stat. Mech. its Appl., vol. 523, pp. 702–713, Jun. 2019, doi:10.1016/j.physa.2019.03.012.

[14]. Chen, Y., Xiong, J., Xu, W., & Zuo, J. (2019). A novel online incremental and decremental learning algorithm based on variable support vector machine. Cluster Computing, 22, 7435-7445.

[15]. Y. Bao and X. Jiang, "An intelligent medicine recommender system framework," 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, 2016, pp. 1383-1388, doi: 10.1109/ICIEA.2016.7603801.

Cite this article as :

K. Saranya, R. Jayashree, A. Ranjitha, "Recommendation of Drugs and Its Substitute Medicines Also Purchasing Using Cosine Similarity Vector", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 3, pp. 28-33, May-June 2023. Available at doi :
https://doi.org/10.32628/IJSRSET231036
Journal URL : https://ijsrset.com/IJSRSET231036