IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

Review on Medicine Recommendation System

Bedade Jayshree Nivrutti, Dr.Mahesh R. Sanghavi, Deepali Vaidya, Waje Ashwini Vijay
Department of Computer Engineering
Late G. N. Sapkal College of Engineering, Nashik, Maharashtra, India
jayshreebedade@gmail.com, sanghavi.mrcoe@snjb.org, deepalivaidya44@gmail.com, ashwiniwaje88@gmail.com

Abstract: *From last two years we see that lot of people are suffering from covid, there is no any medicine for this disease .Some medicine are available but it cannot completely cover covid. Medical section gives lot of doses of medicine to protect people from death. But that extra doses of medicine harmful to body .More and more people are caring about health and medical diagnosis problem .if we see more than 42% people causes with medication error & cause due to expert as per their experience .On internet 60% people search topic related to health information & 35% respondents on diagnosis .Corona virus shows there is shortage of specialty and health care and clinical resources entire medical fraternity distress.so this work intends to present medicine recommendation system that applies the data mining and machine learning to this system. Also apply the map reduce technologies to enlarge the ability of processing big diagnosis of data. In this system also try creation of medicine recommendation system based on graph database.*

Keywords: BOW, TF-IDF, word2vec

I. INTRODUCTION

Now a Days number of coronavirus cases increasing rapidly, lot of countries are facing a shortage of doctors, mostly in rural areas where the quantity of experts is less compared to urban areas. A doctor takes roughly 6 to 12 years to procure the necessary qualifications. Thus, the number of doctors can't be extended quickly in a short time frame. A Telemedicine Framework ought to be energized as far as possible in this difficult time. Clinical bungles are very regular nowadays.

Above 200 thousand persons in China and 100 thousand in the USA are affected every year because of drug mistakes. Above 40% medication, specialists make mistakes while prescribing since specialists create the solution as referenced by their knowledge, which is very restricted. Every patient needs a top level consultation that recognizes wide-based information nearby microscopic organisms, antibacterial medications. Every day a new ideas and research comes up with accompanying more drugs, tests, accessible for clinical staff every day. Accordingly, it turns out to be increasingly challenging for doctors to select which treatment or medications to give to a patient based on symptoms, past medical history.

There are some realities that may lead to these issues: (i) shortage of doctors or medical specialists for serious disease, experience, especially for those inexperienced novices, which are tough to avoid mistakes. For now, most diagnosis case data in hospitals is still reserved untouched and has not stayed used for mining, so that the value behind the data cannot be Explored. Since the mid-1990s, various recommender system techniques have been recommended, and many sorts of recommender system software have been developed recently for a change of applications. Collaborative filtering, content-based, knowledge-based techniques and hybrid recommendation techniques, are used as latest techniques for better recommendation method. Those all having some limitation and need of improvement. Content based has over specialized recommendations, and CF has sparseness, scalability and cold- start problems. To solve all these problems, we apply data mining technologies to recommendation systems and design our own medicine recommender system framework. A recommender system is a customary system that proposes an item to the user, dependent on their advantage and necessity. This medicine recommendation system can go through database system module, Preparation of data module, recommendation system module, evaluation model, data visualization section. Data visualization represents some knowledge which is valuable regarding the analysis case data. In this there is investigation the medicine

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

recommendation algorithms of the Support Vector Machine, BP neural network algorithm and ID3 decision tree algorithm based on the identification records. Lastly, SVM is used for the medicine recommendation model for high accurateness of system, good ability and scalability. In consideration of the safety of the patient, we proposed a mistaken-check mechanism that makes sure the safety and the quality of service. After the completion of this system get medication recommendation with an admirable efficiency, accurateness and scalability. In this, we offer a scientific review of GLRS, by discussing however they extract necessary information from graph-based representations to boost the accuracy, liableness and explain ability of the recommendations. First, we tend to characterize and formalize GLRS, and so summarize and reason the key challenges and main progress during this novel analysis space.

II. LITERATURE SURVEY

Ontology and rule-based drugs recommendation systems: The drug recommendation system Galen OWL[2] relies on the Greek drug guide GALINOS wherever doctors will look for a drug and notice details on the medicine and extra data, like interactions with alternative medicine. The paper describes a system that recommends medicine for a patient supported the malady of the patient, allergies and best- known drug interactions for the medicine within the info. To advocate the most effective fitting drug, rules for medications and interactions area unit hold on within the system that relies on ontologies, ICD-codes and alternative data. This is accessible via the browser.

Data mining and machine learning-based medication recommendation systems: The approach planned by Sun et al.[8] analyzed EMR records to find typical treatment regiments and measures (quantitatively) the effectiveness for those regimens for specific patient cohorts. The authors live the similarity between the treatment records within the EMR, cluster similar ones to treatment regimens supported Map cut back increased Density Peaks based mostly agglomeration, extract semantically pregnant data for the doctor and estimate the treatment outcome for a patient cohort for a typical treatment regime. The results of applying this approach in associate degree empirical study show that the effective rate of the patient will increase moreover because the cure rate.

Recommendation System supported linguistic net technology: The approach planned by Kushwaha[4] et al. describes a drug recommendation system supported linguistics net technology and data processing algorithms. Those 2 strategies were combined to 1st extract linguistics knowledge so apply data processing algorithms on that knowledge. Data processing algorithms were wont to individualize the treatment addicted to the patient's attributes. The system won't suggest medicine that the patient took before or that might act with medicine the patient took before.

A hybrid framework this is proposed in [16] to advocate medicine by ranking is projected in Practitioners creates inquiries and order work tests. Info concerning this patient is entered into the system as a brand new case throughout the method. The system can method the new knowledge and extract patient options. A designation is formed supported the patient's downside. The designation is matched to a selected illness class within the system to see that symptom-drug classifier to use. Patient options within the new case are place into the classifier to predict that drug cluster/clusters to settle on for this patient. Medicine in every cluster is going to be graded by the ranking module to create the ultimate recommendation list.

A Cloud-based: Another drug recommendation system could be a cloud-based platform utilizing numerous algorithms [6].Exploitation the vector service model, the drug character is formatted consistent with the outline of the drug info. Then a k- means rule is applied to cluster medicine. Later on, associate degree analysis exploitation cooperative filtering ends up in recommendations. Finally, tensor decomposition is applied to handle meagerness and large knowledge, shortcomings of cooperative filtering. This multi-step method helps to create associate degree correct recommendation.

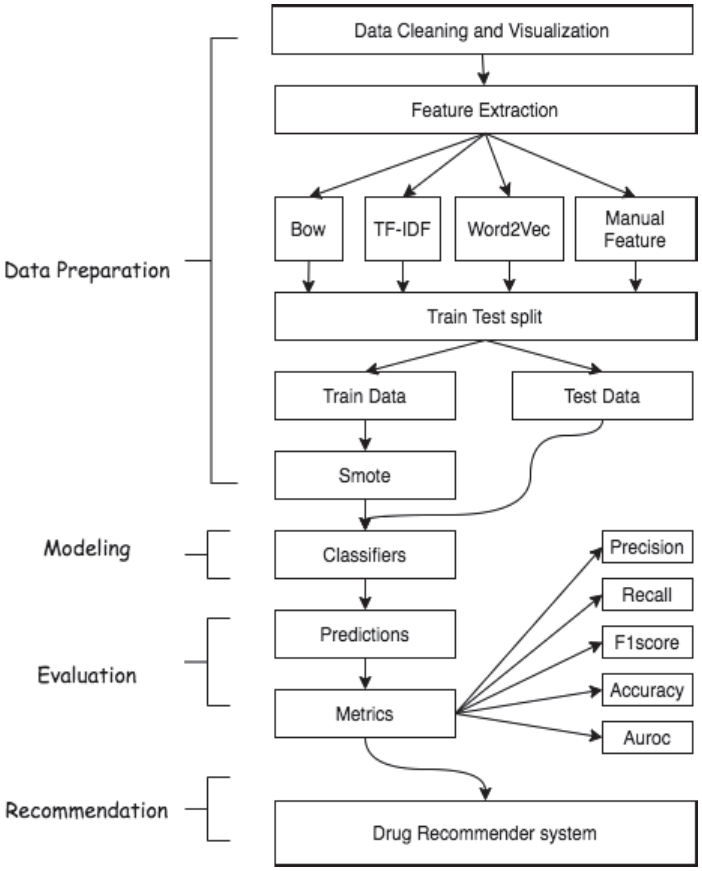
DiaTrack was developed by Medvedeva[7] as a drug recommendation system for sort two polygenic disorders and intends to present doctors a dashboard wherever they'll see similar patient cases and their reaction to a drug or different things. Therefore, the system compares the illness pattern of multiple patients and provides back the leads to a color-coded, straightforward to know graph.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

III. METHODOLOGY

3.1 Data Cleansing and Visualization

Applied customary knowledge preparation techniques like checking null values, duplicate rows, removing gratuitous values, and text from rows during this analysis. After, 1200 null values rows within the conditions column.



3.2 Feature Extraction

After text preprocessing, a correct established of the information needed to create classifiers for sentiment analysis. Machine learning algorithms can't work with text straightforwardly; it ought to be modified over into numerical format. Especially, vectors of numbers. A documented and easy strategy for feature extraction with text data utilized in this analysis is the bag of words (Bow) TF-IDF, Word2Vec. Additionally used some feature engineering techniques to extract features manually from the review column to make another model known as manual feature except for Bow, TF-IDF, and Word2Vec.

1. **Bow:** Bag of words is associate degree algorithmic program utilized in language process chargeable for investigating the amount of times of all the tokens in review or document. A term or token is known as one word (unigram), or any subjective range of words, n- grams during this study, (1,2) n-gram. Outlines however unigrams, diagrams, and trigrams framed from a sentence. The Bow model expertise a major disadvantage, because it considers all the terms while not considering how many terms ar exceptionally ordered within the corpus, that successively build oversized matrix t hat's computationally expensive to coach.
2. **TF-IDF:** TF-IDF maybe widespread weight strategies during which words are offered with weight not count. The principle was to relinquish low importance to the terms that usually seem within the dataset, which means TF-IDF estimates relevancy, not recurrence. Term frequency (TF) is known as the chance of locating a word in

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

a very document.

3. **Word2Vec:** despite the fact that TF and text frequency TFIDF are famed vectorization ways utilized in totally different language preparing task, they disregard the linguistics and grammar likenesses between words. For example, in each TF and TFIDF extraction ways, the words beautiful and pleasant are known as 2 distinctive words in each TF and TF-IDF vectorization like different techniques though they're nearly equivalents. Word2Vec may be a model wont to turn out word embedding. Word embedding's reproduced from large corpora utilizing varied deep learning models Word2Vec takes a massive corpus of text as associate degree input and outputs a vector area, generally composed of hundred dimensions. The elemental thought was to require the linguistics which means of word and prepare vectors of words in vector area with the last word objective that words that share similar sense within the dataset are found on the point of one another in vectors area.
4. **Manual Features:** Feature engineering may be a widespread idea that helps to extend the accuracy of the model. We used fifteen options, that embody useful count, the condition column that is label encoded mistreatment label encoder operate from Scikit library, day, month, year options were developed from date column mistreatment Date Time operate mistreatment pandas. Text blob toolkit [20] was wont to extract the cleansed and uncleaned reviews polarity and else as options at the side of a complete of eight options generated from every of the text reviews

3.3 Train Test Split

Created four datasets mistreatment Bow, TF-IDF, Word2Vec, and manual options. These four datasets were split into 75% of coaching and twenty fifth of testing. whereas cacophonic the information, we set associate degree equal random state to confirm constant set of random numbers generated for the train take a look at split of all four generated datasets.

3.4 Smote

Smote generates the new minority category knowledge by linear interpolation of willy-nilly selected minority instance 'a' together with its k nearest neighbor instance 'b' within the feature area. It displays that there are additional orange points within the non-smote t-SNE projection that represents the bulk category dominance. It additionally shows that there has been associate degree increment in blue points once mistreatment smote that brings out the balance between a majority and minority category that curbs the predominance of the bulk category.

3.5 Classifiers

Distinct machine-learning classification algorithms were wont to build classifier to predict the some provision like a sentiment. provision Regression, Multinomial and another one Naive mathematician, random gradient descent, Linear support vector classifier, Perceptron, and Ridge classifier experimented with bow ,TF-IDF model since they're terribly distributed matrix and applying tree-based classifiers would be terribly long. Applied call tree, Random Forest, LGBM, and CatBoost classifier on Word2Vec and manual options model. A major downside with this dataset is an around 210K review, which takes substantial machine power. We tend to select those machine learning classification different algorithms solely that reduces the coaching time and provides quicker predictions.

3.6 Metrics

The predicted sentiment were measured mistreatment 5 metrics, namely, exactness (Prec), recall (Rec), f1score (F1), accuracy (Acc.) and terrorist organization score.

3.7 Drug Recommender System

After assessing the metrics, all four best-predicted results were picked and joined along to provide the combined prediction. The integrated results were then increased with normalized helpful count to come up with associate degree overall score of drug of a selected condition. the upper the score, the higher is that the drug. The motivation behind the standardization of the helpful count was observing the distribution of helpful count in Fig. 7; one might analyze that the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

distinction among the smallest amount and most extreme is around 1300, appreciable. Moreover, the deviation is gigantic, that is thirty six. the aim behind is that the additional medications people rummage around for, the more people browse the survey in spite of their review is positive or negative, that makes the helpful count high.

IV. CONCLUSION

Design a general medical recommender system A framework for applying data mining techniques to medicine Diagnostics, by database system modules, data preparation module, recommendation model module, model Evaluate models and data visualization modules. We give Concrete realization of each module based on open data set. Experiments are performed to evaluate the model, and finally, SVM was selected as a model for drug recommendation It has high accuracy, good efficiency and scalability in this open data set. Considering patient safety, we also An error checking mechanism to ensure safety is proposed and service quality. In future work, we plan to build our own recommendation model to improve accuracy and the efficiency of the model is further improved. Also plan to apply MapReduce Technology in parallel with our medical recommender system Expanding capabilities to process large diagnostic data.

ACKNOWLEDGEMENT

Inspiration and guidance are invaluable in every aspect of life, especially in the field of education, which I have received from our respected Dr. Mahesh Sanghavi who has guided me in the paper work and gave earnest co-operation whenever required. I also express my sincere gratitude towards her. I am pleased to announce that my paper work would not have been completed without the able guidance and complete support of Dr. Khyati Nirmal who helped me at each and every step in every possible way. He always provided me guidance for latest technology encouragement at every point and took active participation in the achievement of my objective. I also thankful to Dr. N. R. Wankhade my Head of Department. At last, I would like to take this opportunity to say thanks to all staff members, who directly or indirectly encouraged and helped me to complete my work and gives contribution of their valuable time in helping me to achieve success in the paper work.

REFERENCES

[1]. https://www.mohfw.gov.in/pdf/Telemedicine.pdf
[2]. Doulaverakis, C., Nikolaidis, G., Kleontas, A., and Kompatsiaris, I. 2012. GalenOWL: Ontology-based drug recommendations discovery. Journal of Biomedical Semantics 3, 14.
[3]. Doulaverakis, C., Nikolaidis, G., Kleontas, A., and Kompatsiaris, I. 2014. Panacea, a semantic-enabled drug recommendations discovery framework. Journal of Biomedical Semantics 5, 13.
[4]. Kushwaha, N., Goyal, R., Goel, P., Singla, S., and Vyas, O. P. 2014. LOD Cloud Mining for Prognosis Model(Case Study. Native App for Drug Recommender System). AIT 04, 03, 20– 28.
[5]. Zhang, Q., Zhang, G., Lu, J., and Wu, D. 2015. A framework of hybrid recommender system for personalized clinical prescription. 2015 International Conference on Intelligent Systems and Knowledge Engineering.
[6]. Zhang, Y., Zhang, D., Hassan, M. M., Alamri, A., and Peng, L. 2015. CADRE. Cloud-Assisted Drug Recommendation Service for Online Pharmacies. Mobile Netw Appl 20, 3, 348– 355.
[7]. Medvedeva, O., Knox, T., and Paul, J. 2007. DiaTrack. Web- based application for assisted decision-making in treatment of diabetes. Journal of Computing Sciences in Colleges 23, 1, 154– 161.
[8]. 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.