

DRUGS RECOMMENDATION SYSTEM BASED ON PATIENT SENTIMENTS

Malepati Tharunya PG SCHOLAR M.Tech.,(CSE), Dept of CSE,KSRM College of Engineering, Kadapa.

Dr. M. Sreenivasulu Professor, Dept of CSE, KSRM College of Engineering, Kadapa.

Abstract: Most of the people tend to live a long and healthy life, where they are more conscious about their health. But many studies show that almost many people die due to the medical errors caused in terms of taking wrong medicines and these errors are caused by doctors, who prescribe medicines based on their experiences which are quite limited. As machine learning, deep learning and data mining like technologies that are emerging day by day, these technologies can help us to explore the medical history and can reduce medical errors by being doctor friendly. In this paper proposes a medicine recommendation system, which takes the patient review data and performs sentiment analysis on it to find the best medicine for a disease by using Logistic Regression, SVM, Decision Tree, SGD and Navie Bayes model. The paper also discusses the advantages, disadvantages and enhancements that can be incorporated to improve the accuracy.

Keywords: Recommendation system, Sentiment Analysis.

I. INTRODUCTION

This research looks into the feasibility of using sentiment analysis on drug reviews and assesses the efficacy of publicly available sentiment lexicons in the medical field. 1. Sentiment Analysis The extraction and categorization of sentiments conveyed in text data using text analysis techniques is known as sentiment analysis. As evidenced by previous research, sentiment analysis of drug reviews has a lot of promise to help healthcare professionals and companies. Such information can make patients feel safer and have more faith in medical companies. 2. Lexicon-based Sentiment Analysis By matching textual units with opinion terms in sentiment lexicons, lexicon-based sentiment analysis finds features and trends in a given text. Sentiment lexicons are collections of words and phrases that can be used to indicate emotion and are recorded in databases along with their associated sentiment scores. Before being used for sentiment polarity, the sentiment scores in various lexicons are annotated. The lexicon method incorporates sentiment lexicons, which are made up of rules for categorizing words in a text as positive, negative, or neutral. The polarity of a text can be determined using the polarity of sentiment-bearing words in a given text, which is the base of lexicon-based sentiment classification. Typically, a sentiment lexicon is a dictionary that contains the polarity values of words. Tuples of words or phrases are tagged with their corresponding sentiment polarity score in the dictionary. Tuples in the sentiment lexicon database can be expressed in the following way: (word, sentiment polarity score). Each lexicon in a lexicon-driven strategy has a polarity score of positive, negative, or neutral. This paper comprise of 5 sections. Section two provides the related works concerning about these above technologies and intelligent medical system. The section three provides the overview of our proposed medicine recommendation system framework. Section four depicts implementation and results obtained. Section five provides the conclusion and future work.

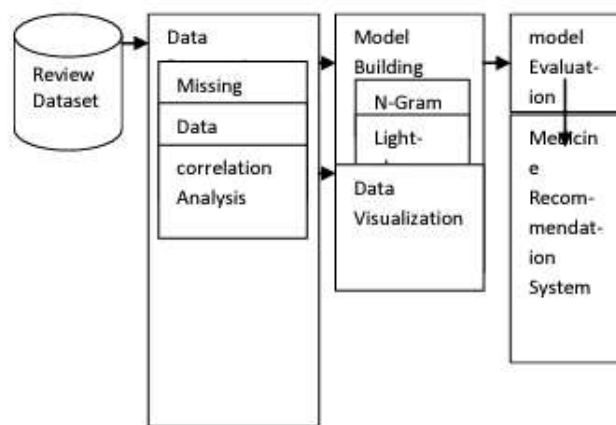


Fig 1: System Architecture

Both healthcare professionals and consumers benefit from drug evaluations since they provide critical medical information. Customers are using online review sites to communicate their ideas and feelings about previously used drugs. A potential buyer, on the other hand, often finds it difficult to read all of the remarks before making a purchasing decision. Another significant difficulty is the reviews' unstructured and textual form, which makes it difficult for users to classify comments into actionable insights. For these reasons, the primary goal of this work is to use text analytics and predictive models within SAS® Enterprise Miner™ to classify the severity of side effects and the effectiveness of prescribed drugs. These models are further validated by using Wit.AI to evaluate performance and generalization. For the goal of identifying the user's opinions (reviews) as good or negative, a dataset that comprises the user's rating and review on the medicine is utilized. The findings can be utilized to provide useful guidelines and references to help potential patients make more educated purchase selections. The findings can also be used to gain a more comprehensive knowledge of the user's perspective on the drug.

II. LITERATURE SURVEY

Recommender frameworks point to supply clients with personalized stock and repair to alter the expanding online information over-burden drawback. Various recommender framework methods are anticipated since the mid1990s, and numerous shapes of recommender framework code were created as of late for a spread of applications. Most of the recommender advances unit of measurement connected to the e-government area[2], e-business area[3], e-commerce/e-shopping area[4], e-learning area[5] and etourism area[6] etc. The medication space incorporates uncommon recommender innovations, and this paper centers on arranging of the medication recommender framework and mining data from therapeutic case knowledge. Through on-line social organizing, the communication is monstrously progressed and totally distinctive intrigued of information is advertised on net essentially at the open pace. The total diverse information must share to chronicle highlights of potential edges and availability of utilities bits of knowledge, things, people practices and items, etc. [7]. One in all the vital areas is that the therapeutic and well being sciences is to think approximately social angles through on-line dialogs, blogs, audits, and on-line overviews, etc. [9]. The health-related substance shared through on-line feedbacks or surveys contains covered up assumption [10] designs that emerges through totally distinctive sources from medical world which offer benefits to the pharmaceutical industry[8]. Amid this, the on-line component is fantastically standard of late for online looking, diverse stock through distinctive websites like on-line buying of drugs at entryway step. Numerous websites and blogs offers clients to rate their stock with their fulfillment and quality of stock, logistics, administrations and criticism etc. , which the clients examines for a particular medicine or on quality of administrations

III. BACKGROUND METHODS

Machine learning is a hot topic in research and industry, with new methodologies developed all the time. The speed and complexity of the field makes keeping up with new techniques difficult even for experts and potentially overwhelming for faster analysis.

Logistic regression

Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. It is one of the supervised learning and is used to

estimate the target object value's possibility. It is a tool to calculate the statistical values and make results on binary output. In the linear method, which is calculated by the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts $P(Y=1)$ as a function of X . Here, y is the linear model's output trained with logistic regression produce value between zero and one.

Naive Bayes

In the Naïve Bayes network, all features are independent. Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles. When there is a change in one feature, it does not affect another. This is suitable for large datasets. The assumption from Conditional independence is that an attribute value is independent of the values, which are from other attribute values in a class. Bayes' Theorem is based on probability theory.

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:

Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.

Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem.

Support Vector Machine (SVM)

SVM is used both for regression and classification tasks. The SVM model represents the data in the space described so that the examples in various categories are divided by a distance as large as possible. That divides sensitive information with the maximum separable space between them and is calculated so that many of the points belong to one group fall on the plane's one side.

IV. PROPOSED SYSTEM

In this segment, the characteristics of the recommender framework are reaching to be specified, additionally the detail of our drugs recommender framework system are aiming to be presented. Recommender framework has gotten to be a profitable investigation field as the advancement of counterfeit brilliantly advances. Not at all like most current recommender frameworks that specialize in e-business, book and pictorial suggestion, our framework points at giving a virtual fully fledged specialist for unpracticed amateurs and patients in abuse right pharmaceutical. Since high accuracy and strength is vital for such an online pharmaceutical recommender framework, in this way we tend to evaluate a few information preparing approaches to induce an genuine trade-off among the precision, productivity and quantifiability. In this proposed System, the framework mainly consists of five modules, as shown in figure 1, which are (i) Database module (ii)Data preparation module (iii)Recommendation model module (iv)Model evaluation model and (v)Data visualization module. Through this process set the topic, preprocess the data to fit the objective, and create various variables to fit model. At the model part, emotion analysis using word dictionary, applying deep learning, etc. were used and then accuracy of both the algorithms is evaluated.

Existing System

The study presents GalenOWL, a semantic-empowered online framework, to help specialists discover details on the medications. The paper depicts a framework that suggests drugs for a patient based on the patient's infection, sensitivities, and drug interactions. For empowering GalenOWL, clinical data and terminology first converted to ontological terms utilizing worldwide standards, such as ICD-10 and UNII, and then correctly combined with the clinical information. Leilei Sun examined large scale treatment records to locate the best treatment prescription for patients. The idea was to use an efficient semantic clustering algorithm estimating the similarities between treatment records. Likewise, the author created a framework to assess the adequacy of the suggested treatment. This structure can prescribe the best treatment regimens to new

patients as per their demographic locations and medical complications. An Electronic Medical Record (EMR) of patients gathered from numerous clinics for testing.

Disadvantages of Existing System

- In the existing work, the system did not implement an exact sentiment analysis for large data sets.
- This system is less performance due to lack Data Classification and Data Fragmentation technique.

Proposed System

A recommender framework is a customary system that proposes an item to the user, dependent on their advantage and necessity. These frameworks employ the customers' surveys to break down their sentiment and suggest a recommendation for their exact need. In the drug recommender system, medicine is offered on a specific condition dependent on patient reviews using sentiment analysis and feature engineering. Sentiment analysis is a progression of strategies, methods, and tools for distinguishing and extracting emotional data, such as opinion and attitudes, from language. On the other hand, Featuring engineering is the process of making more features from the existing ones; it improves the performance of models.

Advantages of Proposed System

- The system is more effective since it presents the proposed algorithm used in natural language processing responsible for counting the number of times of all the tokens in review or document..
- The system has exact sentiment analysis prediction techniques for Data Cleaning and Visualization.

Application Modules

In this segment, the characteristics of the recommender framework are reaching to be specified, additionally the detail of our drugs recommender framework system are aiming to be presented. Recommender framework has gotten to be a profitable investigation field as the advancement of counterfeit brilliantly advances. Not at all like most current recommender frameworks that specialize in e-business, book

and pictorial suggestion, our framework points at giving a virtual fully fledged specialist for unpracticed amateurs and patients in abuse right pharmaceutical. Since high accuracy and strength is vital for such an online pharmaceutical recommender framework, in this way we tend to evaluate a few information preparing approaches to induce an genuine trade-off among the precision, productivity and quantifiability. In this proposed System, the framework mainly consists of five modules

- Database module
- Data preparation module
- Data Exploration
- Data Preprocessing
- Data visualization module.

Through this process set the topic, preprocess the data to fit the objective, and create various variables to fit model. At the model part, emotion analysis using word dictionary, applying deep learning, etc. were used and then accuracy of both the algorithms is evaluated.

Database System Module:

It contains a drug review dataset with attributes like unique Id, drug name, condition(disease of patient), date , useful count, reviews and ratings given by the patients on the drugs.

Data Preparation Module:

It comprises of information investigation and information preprocessing. The real-world information is crude information which can be fragmented, boisterous and messy. Thus, information arrangement is utilized to clean information. it comprises of missing value processing, correlation analysis and removing data redundancy

Data Exploration:

- a) Find unique number of patient ids to check if a patient has written multiple reviews.
- b) Analyze number of drugs per condition by considering condition and number of drugs.

Data preprocessing:

- Find out the number of missing values for all the attributes.
- Find out the set of co-occurring words from the reviews starting from various ML.
- 1-gram: Analyze the text with a single corpus. But it does not classify the emotion well.
- 2-gram: it is hard to classify positive and negative reviews using bi-gram.
- 3-gram: Tri-grams are still unable to classify the positive and negative reviews.
- 4-gram: 4-gram classifies the emotions much better than their grams. Therefore, 4-gram is used to build the deeplearning model.

Visualization module:

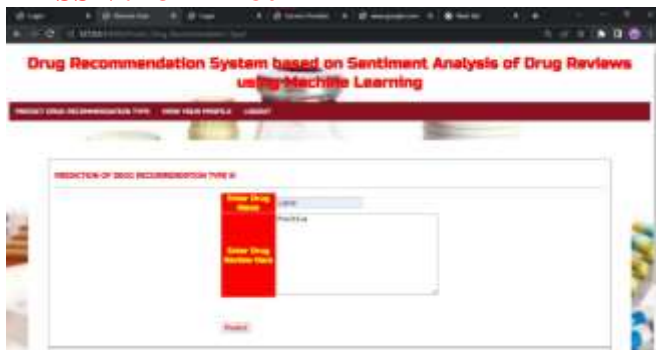
It primarily gives the visualization innovation to show a few important information behind the determination case information

V.RESULTS

A set of experiments carried out on stress analysis data obtained from kaggle.com patients data. The performance evaluation of the system is performing using this dataset. The screenshots of various phases of drugs analysis system are as follows



Screen 1:Service Provider Login Page



Screen 2: Prediction of Drugs



Screen 3 : Chart showing Accuracy using various ML Models



Screen 4: Classification Report Showing Precision, Recall and Score values using Naïve Bayes and SVM



Screen 5: Classification Report Showing Precision, Recall and Score values using Logistic Regression



Screen 6: Classification Report Showing Precision, Recall and Score values using Decision Tree and SGD Classifier

VI.CONCLUSION

The paper has proposed an idea for medicine recommendation system for medical inquiry. This approach is based on four main steps: (i) analysis of review dataset (ii) data preprocessing (iii) model building. iv) Recommending the proper medicine for a particular disease. The proposed system works as a tool for supporting the doctors in their disease diagnosis. As future work efficiency of recommendation system can be increased by including age of the person, demographic information during the training phase. Also the brand and the chemical contents available in the medicine can improve the recommended medicines.

REFERENCES

- [1] S. Fox and M. Duggan. Health online 2013. PewInternet and American Life Project. <http://pewinternet.org/Reports/2013/Healthonline.aspx>, 2013.
- [2] X. Guo, J. Lu, Intelligent e-government services with personalized recommendation techniques, International Journal of Intelligence Systems,2007,401–417
- [3] T. Lee, J. Chun, J. Shim, S.-g. Lee, An ontology-based product recommender system for B2B marketplaces, International Journal of Electronic Commerce, 2006,125–155.

- [4] J.B. Schafer, J. Konstan, J. Riedl, E-commerce recommendation applications, Applications of Data Mining to Electronic Commerce, Springer, US 2001, 115–153.
- [5] O.R. Zaiane, Building a recommender agent for elearning systems, Proceedings of 2002 International Conference on Computers in Education, 2002, 55–59
- [6] T.Hung-Wen,S.VonWun,A personalized restaurant recommender agent for mobile e-service, 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service. EEE, 2004, 259–262.
- [7] Popescu AM, Etzioni O (2005)Extracting product features and opinions from reviews. Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, pp: 339-346
- [8] Mei Q, Ling X, Wondra M, Su H, Zhai C (2007) Topic sentiment mixture: modeling facets and opinions in weblogs. Predictive Modeling of Web Users, pp: 171- 180.
- [9] <https://www.livewell.pk/> [Accessed on: August 2017].
- [10] Aronson AR (2001) Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: AMIA annual symposium proceedings. Washington DC: American Medical Informatics Association, pp: 17-21.