

A SENTIMENTAL ANALYSIS OF DRUG RECOMMENDATION**USING MARITAL LIFE****LIKHITHA THUPAKULA¹, Dr. U. THIRUPALU²****MCA¹, Audisankara College of Engineering & Technology, Gudur, Tirupati, Ap, India****Assoc. Professor² Dept. of MCA Audisankara College of Engineering & Technology, Gudur, Tirupati, Ap, India****ABSTRACT**

Since the corona virus first surfaced, authentic clinical resources are more difficult to get due to a scarcity of specialists, healthcare professionals, appropriate equipment, and medications, among other factors. There are many people dying because the medical profession as a whole is in trouble. Due to lack of availability, people began self-medicating without proper advice, which made the health condition worse than usual. Machine learning has recently shown useful in many areas, and creative work for automation is increasing. This essay aims to propose a medication recommendation method that might significantly lessen the workload of specialists. In this study, we develop a medication recommendation system that uses customer reviews to predict the sentiment using a variety of vectorization techniques like Bow, TF-IDF, Word2Vec, and manual feature analysis, which can assist in recommending the best medication for a given disease by various classification algorithms. Predictive feelings were assessed using precision, recall, f1score, accuracy, and AUC score. The outcomes demonstrate that classifier LinearSVC employing TF-IDF vectorization performs better than the other models with 93% accuracy.

1. INTRODUCTION

The world is experiencing a doctor shortage as a result of the corona virus, especially in rural areas where there are fewer specialists than in urban areas. The training required to become a doctor often lasts 6 to 12 years. As a result, it is unable to hire more doctors fast. In this trying time, it is important to keep a Telemedicine framework powered. Clinical blunders are very regular now a days. Over 200 thousand individuals in China and 100 thousand in the USA are affected every year because of prescription mistakes. Over 40% medical specialists make mistakes while prescribing since specialists compose the solution as referenced by their knowledge, which is very restricted. Choosing the to p-level medication is significant for patients who need specialists that know wide based information about micro scope organisms, antibacterial medications, and patients. Everyday a new study comes up with accompanying more drugs, tests, accessible for clinical staff every day.

past clinical history with the exponential development of the web and the web-based business industry, item reviews have become an imperative and integral factor for acquiring items worldwide. Individuals worldwide become adjusted to analyze reviews and websites first before settling on a choice to buy a thing. While most of past exploration zeroed in on rating expectation and proposals on the E-Commerce field, the territory of medical care or clinical the raps have been in frequently taken care of. There has been an expansion in the number of individuals worried about their well-being and finding a diagnosis online. As demonstrated in a Pew American Research center survey directed in 2013, roughly 60% of grown-up searched online for health-related subjects, and around 35% of users looked for diagnosing health conditions on the web. A medication recommender frame work is truly vital with the goal that it can vital with the goal that it can assist specialists and help patients to build their knowledge of drugs on specific health conditions. A recommender framework is a customary system that proposes an item to the user, dependent on their advantage and necessity. These frame works 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 engineering of features and analysis. Sentiment analysis is a continuum of techniques, methods, and tools for identifying and extracting emotional information from language, such as attitudes and opinions. The technique of "feathering engineering," on the other hand, entails adding new features to the ones that already exist in order to enhance models' functionality. Five sections make up this examination work: A brief explanation of the need for this research is given in the introduction section. The related works section provides a concise overview of prior studies on this topic. The applied model outcomes are evaluated by the Result segment using a variety of metrics, followed by the Discussion part that discusses the framework's shortcomings and the conclusion section.

2.LITERATURE SURVEY

Application of machine learning and deep learning has increased due to the rapid growth of AI. the arXiv:2104.01113v2 [cs.LG] techniques for recommender frameworks, 5 April 2021. These days, recommender systems are widely used in the restaurant, e-commerce, travel, and other sectors. Due to the fact that medication reviews are much more difficult to analyze because they contain clinical terminology like names of infections, reactions, and synthetic names that are used in the production of the drug, there are sadly few studies in the field of drug proposal framework utilizing sentiment analysis.

In order to aid specialists in learning more about the drugs, the study offers Galen OWL, a semantically enhanced web platform. In the paper, a framework that recommends medications for a patient based on the patient's infection, sensitivities, and drug interactions is presented. For the purpose of empowering Galen OWL, clinical data and terminology were first converted to ontological terms using international standards like ICD-10 and UNII, and after that, they were correctly combined with the clinical information. Leilei Sun examined extensive treatment records to find the best treatment prescription for patients. The plan was to estimate the similarities between treatment data using a semantic clustering technique that is effective.

This structure can prescribe the best treatment regimens to new patients as per their demographic locations and medical complicates. An Electronic Medical Record (EMR) of patients gathered from numerous clinics for testing. The result shows that this framework improves the curate. In this research, multilingual sentiment analysis was performed using Naive Bayes and Recurrent Neural Network(RNN). Google translator API was used to convert multilingual tweets into the English language. The results exhibit that RNN with 95.34% out performed NaïveBayes, 77.21%.

3.SYSTEM ANALYSIS

The main concept behind this drug recommendation system is to build a recommendation that helps the patient by

3.3 PROPOSED SYSTEM

According to the user's benefit and need, a recommender framework is a common system that makes an item recommendation. These frameworks analyses customer surveys' results to deconstruct respondents' attitude and offer recommendations tailored to each individual's particular requirements. Using sentiment

recommending suitable medicine for their disease. This system offers the medicine commendation system that helps the patient by recommending suitable medicine for their disease. This system offers the medicine recommendation by recommending the drug based on the specific condition dependent on the patient reviews.

3.1 EXISTINGSYSTEM

Recommender frameworks point to supply clients with personalized stock and repair to alter the expanding online information over-burdened raw back. Various recommender framework methods are anticipated since the mid1990s, and numerous shapes of recommender framework code were created as of late for as do applications.

The health- related substance shared through on-line feedback or surveys contains covered up assumption designs that emerges through totally distinctive sources from medical world which offer benefits to the pharmaceutical industry. 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 entry way step. Numerous websites and blogs offer clients to rate their stock with their fulfillment and quality of stock, logistics, administrations and criticism etc., which the clients examine for a particular medicine or on quality of administration.

3.2 DISADVANTAGES OF EXISTING SYSTEM

In the existing work, the system did not implement an exact sentiment analysis for large data sets. Analysis is a series of approaches, techniques, and instruments for identifying and separating emotional information from language, such as attitudes and opinions. Feature engineering, on the other hand, is a method for enhancing models' functionality by adding new features to the ones that already exist.

analysis and feature engineering, the drug recommender system offers medications under certain conditions that

3.4 ADVANTAGES OF PROPOSED SYSTEM

- The system is more efficient since it displays the proposed natural language processing algorithm that

counts the frequency of each token in a document or review.

•For data cleaning and visualization, the system offers precise sentiment analysis prediction tools.

4.SYSTEMARCHITECTURE

Architecture defines the components, modules interfaces and data for a system to satisfy specified requirements. One should see as the applications of the systems theory to product development. System architecture bridges the gap between the problem domain and the existing system in a manage able way. It is the phase where the SRS document is converted in to a format that can be

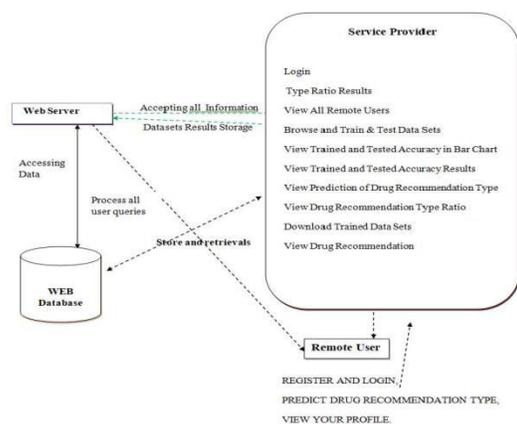


Figure1: Architecture Diagram

5.MODULEDESCRIPTION

5.1 MODULES

5.1.1 SERVICEPROVIDER

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse and Train & Test Data Sets , View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Drug Recommendation Type, View Drug Recommendation ,Type Ratio Download Trained Data Sets, View Drug Recommendation, Type Ratio Results, View All Remote Users.

5.1.2 VIEWANDAUTHORIZEUSER

Details such as, username, email address and admin authorize the users. In this module, the admin can view the list of users who all registered. In this the admin can view the users.

5.1.3 REMOTEUSER

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their data Implemented and decide show the system will operate. In this phase, the complex activity of system development is divided into several smaller sub activities, which coordinate with each other to achieve the marriage in objective of system development. stored to the database. After registration successful, he has to login by using authorized username and password. Once Login is successful user will do some operations like register and Login, predict drug recommendation type, view your profile.

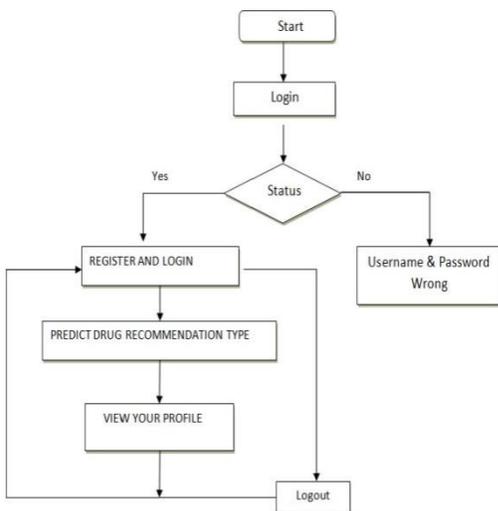
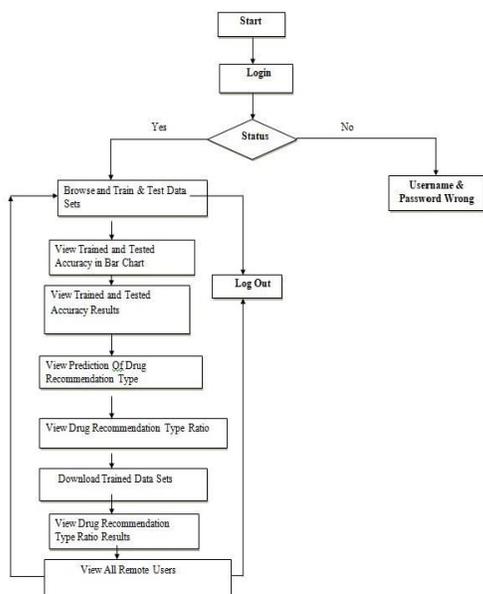


Figure3: Flow chart of remote user

5.2 METHODOLOGIES

The data set taken in this research contains six attributes, name of drug used, review of a patient, condition of a patient, useful count which suggest the number of individuals who found the review helpful, date of review entry, and a 10-star patient rating determining overall patient contentment. The proposed model used to build a medicine.

6.1 DATA CLEANING AND VISUALIZATION

Applied standard Data preparation techniques like checking null values, duplicate rows, removing unnecessary values, and text from rows in this research. We make sure that a unique it should be unique to remove delicacy.

FEATURE EXTRACTION

After text preprocessing, a proper set up of the data required to build classifiers for sentiment analysis. Machine learning algorithms can't work with text straight forwardly; it should be changed over into numerical format. In particular, vectors of numbers. A well-known and straight forward strategy for feature extraction with text information used in this research is the bag of words (Bow), TF-IDF, Word2Vec. Also used some feature engineering techniques to extract features manually from the review column to create another model called manual feature.

6.1.1 BOW

language processing responsible for counting the number of times of all the tokens in review or document. A term or token can be called one word(unigram), or any subjective number of words, n-grams. In this study, (1,2) n-gram range is chosen. The Bow model experience a significant drawback, as it considers all the terms without contemplating how a few terms are exceptionally successive the corpus, which in turn build a large matrix that is computationally expensive to train.

6.1.2 TF-IDF

TF-IDF is a popular weighting strategy in which words are offered with weight not count. The principle was to give low importance to the terms

that often appear in the dataset, which implies TF-IDF estimates relevance, not are currency Term frequency (TF) can be called the likelihood of locating a word in a document. Inverse document frequency (IDF) is the opposite of the number of times a specific term showed up in the whole corpus. It catches how a specific term is documentspecific. TF-IDF is the multiplication of TF with IDF, suggesting how vital and relevant a word is in the document. Like Bow, the selected n-gram range for TF-IDF in this work is (1,2).

6.1.3 WORD2VEC

Even though TF and TF-IDF are famous vectorization methods used in different natural language preparing tasks, they disregard the semantic and syntactic likenesses between words. For instance, in both TF and TF-IDF extraction methods, the words lovely and delightful are called two unique words in both TF and TF-IDF vectorization techniques although they are almost equivalents. Word2Vec is a model used to produce word embedding. Word embedding reproduced from gigantic corpora utilizing various deep learning models. Word2Vec takes an enormous corpus of text as input and outputs a vector space, generally composed of hundred dimensions. The fundamental thought was to take the semantic meaning of words and arrange vectors of words in vector space with the ultimate objective that words that share similar sense in the dataset are found close to one another in vector space.

6.1.4 MANUALFEATURE

Feature engineering is a popular concept which helps to increase the accuracy of the model. We used fifteen features, which include useful count, the condition column which is label encoded using label encoder function from Scikitlibrary, day, month, year features were developed from date column using Date Time function using pandas. Text blob tool kit was used to extract the cleaned and uncleaned reviews polarity and added as features along with a total of 8 features generated from each of the text reviews.

6.2 TRAINTESTSPLIT

We created four data sets using Bow, TF-IDF, Word2Vec, and manual features. These four datasets were split into 75% of training and 25% of testing. While splitting the data, we set an equal random state to ensure the same set of random

numbers generated for the train test split of all four generated datasets.

6.3 SMOTE

After the Train Test split, only the training data under gone a synthetic minor over-sampling

technique (Smote) to prevent the class imbalance problem. Smote is an oversampling technique that synthesized new data from existing data. Smote generates the new minority class data by linear interpolation of randomly selected minority instance 'a' in combination with its k-nearest neighbor instance 'b' in the feature space.

6.4 CLASSIFIERS

Distinct machine-learning classification algorithms were used to build a classifier to predict the sentiment. Logistic Regression, Multinomial NaiveBayes, Stochastic gradient descent, Linear support vector classifier, Perceptron, and Ridge classifier experimented with the Bow, TF-IDF model since they are very sparse matrix and applying tree-based classifiers would be very time-consuming. Applied Decision tree, Random Forest, LGBM, and Cat Boost classifier on Word2Vec and manual features model. We selected those machine learning classification algorithms only that reduces the training time and give faster predictions.

6.5 METRICS

The predicted sentiment was measured using five metrics, namely, precision (Perc), recall (Rec), f1score(F1), accuracy (Acc.) and AUC score. Let the letter be: T_p = True positive or occurrences where model predicted the positive sentiment truly, T_n = True negative or occurrences where model predicted the negative class truly, F_p = False positive or occurrences where model predicted the positive class falsely, F_n = False negative or occurrences where model predicted the negative class falsely.

$$\text{Precision} = \frac{T_p}{T_p + F_p}$$

$$\text{Recall} = \frac{T_p}{T_p + F_n}$$

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

$$\text{F1score} = 2 \cdot \left(\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \right)$$

Area under curve (Auc) score helps distinguish a classifier's capacity to compare classes and utilized as a review of the region operating curve(roc)

Curve. Occur visualizes the relationship between true positive rate (Tpr) and false positive rate (Fpr) across various thresholds.

6.6 DRUG RECOMMENDER SYSTEM

After assessing the metrics, all four best-predicted results were picked and joined together to produce the combined prediction. The merged results were the n-multiplied with normalized useful count to generate an overall score of drugs of a particular condition. The higher the score, the better is the drug. The motivation behind the standardization of the useful count was looking at the distribution of useful count. The purpose behind is that the more medications individuals search for, the more individuals read the survey regardless of their review is positive or negative, which makes the useful count high. So while building the recommender system, we normalized useful count by conditions.

SCREENSHOTS



Figure:LoginPage



Figure:Login Servic eProvider



Figure: Registering details



Figure: View all remove users



Figure: Accuracy of algorithm



Figure: Prediction of drug recommendation type

Sublingual	"I had some bad reactions when I was on it, but I was on it for a few months so I was able to work. I had this was going to work. So, I have been on Trifluoro for six months now with Metformin. When I had the first month period the diarrhea, gas, and bloating were really bad. I was having some of those side effects. However, I still have a 40% loss of weight and I am feeling better. I was on it for a few months and I was able to work. I was on it for a few months and I was able to work. I was on it for a few months and I was able to work." Positive
Oral	"I have been on Trifluoro for about a week now, so we will see what this does. I was diagnosed with stage 2 chronic kidney disease (CKD 2) so I am watching diet closely." Negative
Injectable	"I have been on Trifluoro for about a week now, so we will see what this does. I was diagnosed with stage 2 chronic kidney disease (CKD 2) so I am watching diet closely." Negative
Oral	"I was on Trifluoro for about a week now, so we will see what this does. I was diagnosed with stage 2 chronic kidney disease (CKD 2) so I am watching diet closely." Positive
Sublingual	"I was on Trifluoro for about a week now, so we will see what this does. I was diagnosed with stage 2 chronic kidney disease (CKD 2) so I am watching diet closely." Negative

Figure: Drug reviews

CONCLUSION

Reviews are becoming an integral part of our daily lives; whether go for shopping, purchase something online or go to some restaurant, we first check the reviews to make the right decisions. Motivated by this, in this research sentiment analysis of drug

Reviews was studied to build a recommender system using different types of machine learning classifiers, such as Logistic Regression, Perceptron, Multinomial NaiveBayes, Ridge classifier Stochastic gradient descent, Linear SVC, applied on Bow, TF-IDF, and classifiers such as Decision Tree, Random Forest, Lg bm, and Cat boost were applied on Word2Vec and Manual features method. We evaluated the using five different metrics, precision, recall f1score, accuracy, and AUC score, which reveal that the Linear SVC on TF-ID Four performs all other models with 93% accuracy. On the other hand, the Decision tree classifier on Word2Vec showed the worst performance by achieving only 78%accuracy. We added best-predicted emotion values from each method, Perceptron on Bow (91%), Linear SVC on TF-IDF (93%), LGB Mon Word2Vec (91%), Random Forest on manual features (88%), and multiply them by the normalized useful Count to get the overall score of the drug by condition to build are commander system. Future work in involves comparison of different oversampling techniques using different values of n-grams, and optimization of algorithms to improve the performance of their commander system.

REFERENCES

1. WittichCM, BurkleCM, LanierWL. Medication errors: an over view for clinicians. Mayo Clin Proc. 2014Aug;89(8):1116-25.
2. CHEN, M. R., & WANG, H. F. (2013). Thereason and prevention ofhospital medication errors.PracticalJournal ofClinical Medicine
3. Fox,Susannah&Duggan,Maeve.(2012).Health Online 2013. Pew Research Internet Project Report.
4. ShimadaK,TakadaH,MitsuyamaS,etal Drug-recommendation system for patients with infectious diseases. AMIAAnnu SympProc.2005;2005:1112.
5. Zhang, Yin & Jin,Rong& Zhou,ZhiHua.(2010). Understanding bag of words model: A statistical framework. International Journal of Machine Learning and Cybernetics.1.43-52.10.1007/s13042-010-0001-0.
6. J. Ramos et al., "Using tf-idf to determine word relevance in document queries," in Proceedings of the first instructional conference on machine learning, vol.242,pp.133–142Piscataway,NJ, 2003.
7. YoavGoldbergandOmerLevy.word2vecExplained : deriving Mikolovetal.'snegative-samplingwordembeddingmethod,2014;arXiv:1402.3722.
8. DanushkaBollegala, TakanoriMaeharaandKenichiKawarabayashi.,2015;arXiv:1505.07184.
9. van der Maaten, Laurens & Hinton, Geoffrey.(2008).Vializingdatausingt-SNE.JournalofMachineLearningResearch.9.2579-2605.
10. N.V.Chawla,K.W.Bowyer,L.O.Halland W.P.Kegelmeye.SMOTE:SyntheticMinorityOver-samplingTechnique,2011,JournalOfArtificial Intelligence Research, Volume 16, pages321-357, 2002;arXiv:1106.1813. DOI:10.1613/jair.953.
11. Powers, David & Ailab,. (2011). Evaluation:Fromprecision,recallandF-measuretoROC,informedness,markedness&correlation.J.Mach.Learn.Technol.2.2229-3981. 10.9735/2229-3

AUTHOR'S :

Likhitha Thupakula has pursuing her MCA from Audisankara College of Engineering & Technology (AUTONOMOUS) Gudur, affiliated to JNTUA in 2022. Andhra Pradesh, India.



Dr. U. Thirupalu received M. Tech. degree in CSE from Nagarjuna university, Guntur, Andhra Pradesh, India in 2010 and also received Ph.D. degree from Sri Venkateshwara University, Tirupati, Andhra Pradesh, India, in July 2022. He has guided P.G and U.G students in research area including Cryptography and Network Security, Machine Learning etc., At present he is working as an Associate Professor in Audisankara College of Engineering & Technology (AUTONOMOUS), Gudur. Tirupati (Dt), Andhra Pradesh, India.