Drug Recommendation System Using RNN-BiLSTM Algorithm

[1]Mr.B.V.Praveen Kumar, [2]Ch.K.N.M.Dattathreya, [3]N.Sai Sekhar, [4]V.Akanksha, [5]R.Ramyasri,[1]Assistant Professor, Department of COMPUTER SCIENCE AND ENGINEERING,USHA RAMA COLLEGE OF ENGINEERING AND TECHNOLOGY,Telaprolu ,AP,India.

cse.praveen@usharama.in

[2,3,4,5]Department of COMPUTER SCIENCE AND ENGINEERING,USHA RAMA COLLEGE OF ENGINEERING AND TECHNOLOGY,Telaprolu,AP,India

---

Abstract:Since coronavirus has shown up, inaccessibility of legitimate clinical resources is at its peak, like the shortage of specialists and healthcare workers, lack of proper equipment and medicines etc. The entire medical alliance is in distress, which results in numerous individual's demise. Due to inaccesibility, individuals started taking medication independently without appropriate consultation, making the health condition worse than usual.As of late, machine learning has been valuable in numerous applications, and there is an increase in innovative work for automation.This paper intends to present a drug recommender system that can drastically reduce specialists heap.In this research, we build a medicine recommendation system that uses patient reviews to predict the sentiment using various vectorization processes like Bow, TF-IDF, Word2Vec, and Manual Feature Analysis, which can help recommend the top drug for a given disease by different classification algorithms.

Keywords:Recommendation,Machine Learning, NLP,Smote, Bow, TF-IDF, Word2Vec, Sentiment analysis,RNN,LSTM.

I. INTRODUCTION

With the number of corona virus cases growing exponentially, the nations are facing a shortage of doctors, particularly in rural areas where the quantity of specialists 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 expanded quickly in a short time frame.

A Tele-medicine framework ought to be energized as far as possible in this difficult time [1-2]. Clinical blunders are very regular nowadays. Over 200 thousand individuals in China and 100 thousand in the USA are affected every year because of prescription mistakes. Over 40% medicine, specialists make mistakes while prescribing since specialists compose the solution as referenced by their knowledge, which is very restricted[3-4][5]. The isolation can be improved by placing metal strips between

elements [6-7]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. [8-11].

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.

II. LITERATURE SURVEY

With a sharp increment in AI advancement, there has been an exertion in applying machine learning and deep learning strategies to recommender frameworks. These days, recommender frameworks are very regular in the travel industry, e-commerce, restaurant, and so forth. Unfortunately, there are a limited number of studies available in the field of drug proposal framework utilizing sentiment analysis on the grounds that the medication reviews are substantially more intricate to analyze as it incorporates clinical wordings like infection names, reactions, a synthetic names that used in the production of the drug [8].

The study [8] 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 [9] 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. The result shows that this framework improves the cure rate. In this research [11], 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% outperformed Naive Bayes, 77.21%.

The study is based on the fact that the recommended drug should depend upon the patient's capacity. For example, if the patient's immunity is low, at that point, reliable medicines ought to be recommended. Proposed a risk level classification method to identify the patient's immunity. For example, in excess of 60 risk factors, hypertension, liquor addiction, and so forth

have been adopted, which decide the patient's capacity to shield himself from infection. A web-based prototype system was also created, which uses a decision support system that helps doctors select first-line drugs.

Xiaohong Jiang [10] examined three distinct al-gorithms, decision tree algorithm, support vector machine (SVM), and backpropagation neural network on treatment data. SVM was picked for the medication proposal module as it performed truly well in each of the three unique bound- aries - model exactness, model proficiency, model versatility. Additionally, proposed the mistake check system to ensure analysis, precision and administration quality.

Mohammad Mehedi Hassan et al. [11] developed a cloud- assisted drug proposal (CADRE). As per patients' side effects, CADRE can suggest drugs with top-N related prescriptions. This proposed framework was initially founded on collabora- tive filtering techniques in which the medications are initially bunched into clusters as indicated by the functional description.data. However, after considering its weaknesses like com- putationally costly, cold start, and information sparsity, the model is shifted to a cloud-helped approach using tensor decomposition for advancing the quality of experience of medication suggestion.

Considering the significance of hashtags in sentiment anal- ysis, Jiugang Li et al. [7] constructed a hashtag recommender framework that utilizes the skip-gram model and applied con- volutional neural networks (CNN) to learn semantic sentence vectors. These vectors use the features to classify hashtags using LSTM RNN. Results depict that this model beats the conventional models like SVM, Standard RNN. This explo- ration depends on the fact that it was undergoing regular AI methods like SVM and collaborative filtering techniques; the semantic features get lost, which has a vital influence in getting a decent expectation

III.METHODOLOGY

This paper proposes a recommendation system, which takes health condition as the input and recommends the drugs based on the reviews. This system uses a hybrid RNN stacked with bi-directional LSTM model and a gradient boosting framework.

A.Recurrent Neural Network

Recurrent Neural Network (RNN) is a Neural Network, which uses backward propagation, where the result obtained from the step before is fed to the current step as input. In traditional neural network all the input and outputs are independent, but in cases such as when it is important to predict the next word of a sentence, the preceding words are required and, therefore, it is necessary to remember the previous words. This led to the existence of RNN, hence RNN with help of hidden layer was able to solve this issue. RNN's principal and most significant feature is the hidden state, the hidden state has some memory of the sequence.

B. Bi-directional Long Short Term Memory

Bi-directional Long Short Term Memory (BiLSTM) structure allows networks to have information about the sequence both

backward and forward at all times. Using bidirectional LSTM, your inputs will run in two ways: one from past to future and the other from future to past and what varies from unidirectional is that in the LSTM that operates backwards, you are able to preserve knowledge from the future and use the two hidden states together to maintain details from the past and the future at any point in time.

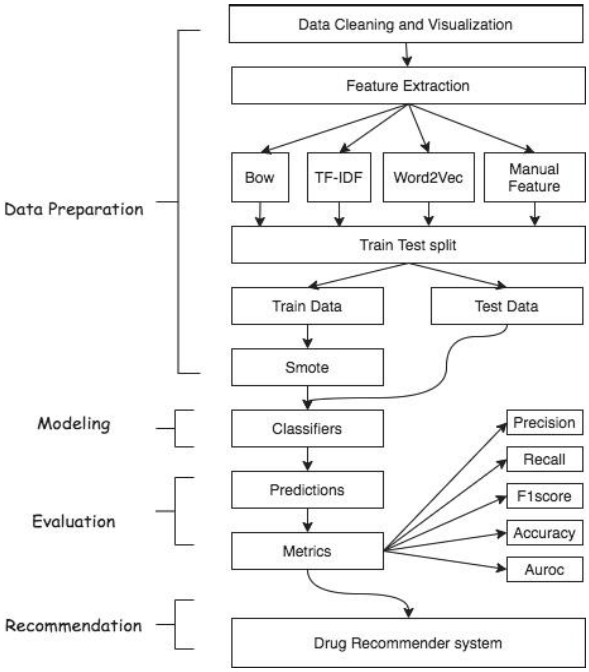

Fig1:Flow chart of proposed model

C. Dataset Gathering

The dataset used in this paper was found at UCI Machine Learning Repository, named as 'Drug Review Dataset (Drugs.com) Data Set'. This dataset is made by web crawling the information from Drugs.com. The dataset contains 215063 instances. As shown in fig2, the dataset contains six attributes, namely drugName which states the name of the drug, condition which states the name of the condition associated with that drug, review which gives the reviews of patients who has used the drug for that specific condition, rating which is the score from 10 given by the patients to the drugs, date which gives the date when the review was posted and useful Count which gives the number of users who have found that review useful. The dataset is divided into training and testing set.

	Column Name	Explanation
0	Id	Drug Id No
1	drugName	Name Of the Drug
2	condition	Functionality of the Drug
3	review	Review of the Drug
4	rating	Rating for Drug
5	date	Drug Manufactured Date
6	usefulCount	Drug Useful Count

Fig2:Dataset attributes

D.Exploratory Data Analysis

It is important to analyse the different attributes and understand their features. First, we analyzed the distribution of ratings across the reviews. From fig.3 we can understand how many reviews were rated above 5 and below 5.

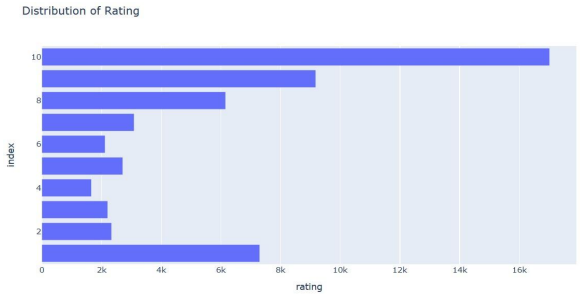

Fig3:Distribution of rating

As shown in fig.4 we have found the top 30 conditions for which the patient has given reviews. We can understand that most of the reviews are about birth control followed by depression with the second highest reviews. The health conditions like pain, anxiety and acne has almost same count. From fig. 4 we gather the information about the most popular drugs.

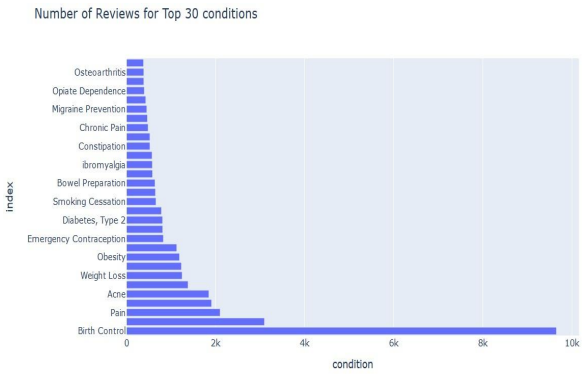

Fig 4 Bar plot of Top 30 Conditions

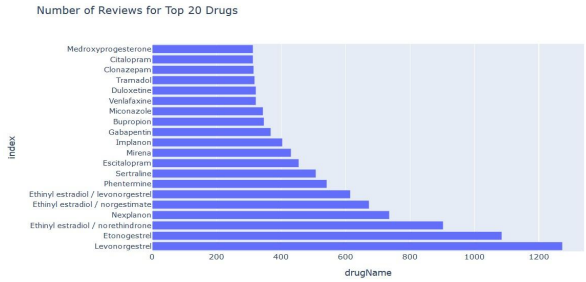

Fig. 5. Bar plot of Top 20 Drugs

E.*Data Pre-processing*

The drug reviews are cleaned by eliminating all the whitespace, converting to lower case letters, collecting the stopwords and all the other general processing techniques. To improve the accuracy and reduce the risk of overfitting, feature extraction and Bag of Words technique is used. Stemming is the next process used which converts the words to its base form, so that words of various forms can be treated as same as their root word [10].

i.Feature Extraction

It is used to reduce the number of features in dataset by creating new features from existing ones. The new reduced feature is used to summarise the most of the information present in the previous ones [10].

ii.Bag Of Words(BOW)

We cannot send our text directly into any algorithm. It is used to pre-process the text by translating it into a bag of words that holds a count of the cumulative occurrences of the most commonly used words [10].

iii.Train Test Split

We created four datasets 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.

iv. SMOTE

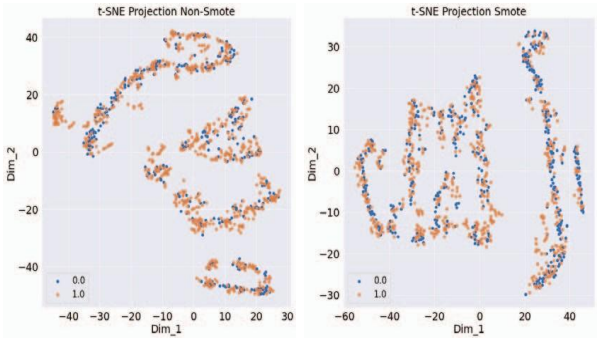

Fig. 6. t-SNE subplot before and after Smote using 1000 training samples

While building the recommender system,we normalized useful count by conditions.

IV.RESULTS

A.Analyzing Different Algorithms.

We have analyzed different algorithms on the dataset tounderstand which one gives the better accuracy. The algorithms used are Naive Bayes, Random Forest, Linear SVC, Logistic Regression and RNN-BiLSTM. The algorithm giving the best accuracy was selected to train the dataset. The accuracy of the algorithm are given in the table1.The best accuracy is given in bold.

Algorithm	Accuracy
Multinomial Naive Bayes	0.75354
Random Forest	0.82926
Linear SVC	0.58199
Logistic Regression	0.63778
RNN-BiLSTM	**0.83906**

Table1:comparing accuracy of different alogrithms

B.Using RNN- BiLSTM Model

The dataset is trained for 10 epochs with the batch size of 64 using this model In recurrent neural network, layers which receive lower gradient stops learning. So, the neural networkcannot process the long sequence and are short term.The long short term memory algorithm processes the entire sequence of data as they have gates which regulates the flowof information. The fig.7 depicts the user's giving review about particular drug and fig.8 shows the prediction result of the users review after performig exact sentiment analysis of the user's review. This model gave the accuracy of 83%.

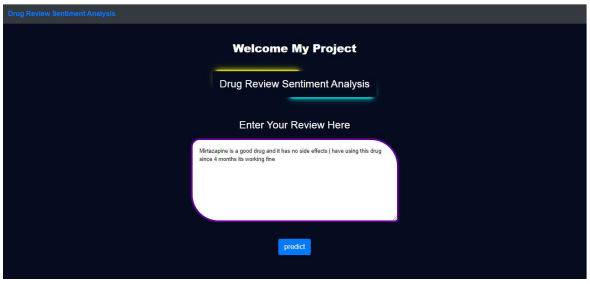

Fig 7 user's review for prediction

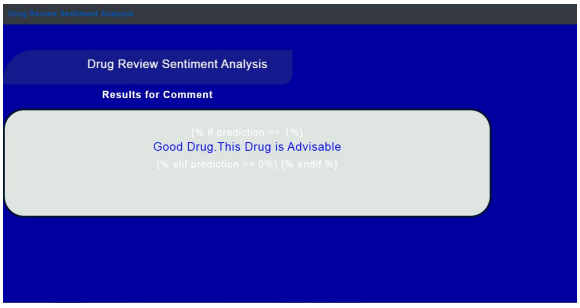

Fig 8: prediction of user's review

V.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,With the advent of immense technological developments, especially the world wide web , individuals have found their ability to express opinions on a variety of products available in market. One such field is reviewing drugs for a medical conditions. With many people relying on these reviews , extracting information from these reviews helps to identify whether a particuular drug is proving to be beneficial as well as discover the aspect that might anger clients.

In this paper we have proposed a drug recommendation system that helps to recommend the medications based on the reviews gained from their users. It is useful to understand the best possible medication for a condition and also helps in drug repurposing..

We have used RNN BiLSTM algorithm to recommend medicines which provides an accuracy of 83%. As a part of our future work. We would also like to use more granular user information such as user age, gender and treatment span to further improve outcomes and improve insights..

REFERENCES

[1] Telemedicine,https://www.mohfw.gov.in/pdf/Telemedicine.pdf
[2] Wittich CM, Burkle CM, Lanier WL. Medication errors: an overview for clinicians. Mayo Clin Proc. 2014 Aug;89(8):1116-25.
[3] CHEN, M. R., & WANG, H. F. (2013). The reason and prevention of hospital medication errors. Practical Journal of Clinical Medicine,
[4] Dataset,https://archive.ics.uci.edu/ml/datasets/Drug%2BReview%2BDataset%2B%2528Drugs.com%2529#
[5] Fox, Susannah, and Maeve Duggan. "Health onlie."URL:http://pewinternet.org/Reports/2013/Health-online.aspx.
[6] Bartlett JG, Dowell SF, Mandell LA, File TM Jr, Musher DM, Fine MJ. Practice guidelines for the management of community-acquired pneumonia in adults. Infectious Diseases Society of America. Clin Infect Dis. 2000 Aug;31(2):347-82. doi: 10.1086/313954. Epub 2000 Sep 7. PMID: 10987697; PMCID: PMC7109923.
[7] Fox, ysis, Jiugang Li & Duggan, Maeve. (2012). Health Online 2013. Pew Research Internet Project Report. [8] 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] Doulaverakis, C., Nikolaidis, G., Kleontas, A. et al. GalenOWL: Ontology-based drug recommendations discovery. J Biomed Semant 3, 14 (2012). https://doi.org/10.1186/2041-1480-3-14
[9] Leilei Sun, Chuanren Liu, Chonghui Guo, Hui Xiong, and Yanming Xie. 2016. Data-driven Automatic Treatment Regimen Development and Recommendation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 1865–1874. DOI:https://doi.org/10.1145/2939672.2939866
[10] Xiaohong Jiang, K. Gupta and N. Kumar, "Sentiment Analysis of Multilingual Twitter Data using Natural Language Processing," 2018 8th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2018, pp. 208-212, doi: 10.1109/CSNT.2018.8820254.
[11] Hassan, Takada H, Mitsuyama S, et al. Drug-recommendation system for patients with infectious diseases. AMIA Annu Symp Proc. 2005;2005:1112.