ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4

Journal of Coastal Life Medicine

Drug Prescribing System Using Patient Reviews Based on Sentimental Analysis

Received: 19 February 2023, **Revised:** 21 March 2023, **Accepted:** 25 April 2023

[a]Dr. P. Sivakumar, [b]N. Nanthini, [b]S. Suruthi, [b]T. Veronica

[a]Professor, Department of Information Technology, Manakula Vinayagar Institute of Technology, Puducherry, India.
[b]Department of Information Technology, Manakula Vinayagar Institute of Technology, Puducherry, India.
Email: hodit@mvit.edu.in, nanthininanmaran1422@gmail.com, suruthisankar32@gmail.com, veronicathaiya6@gmail.com

Keywords:
Deep Learning, Sentimental Analysis, Random Forest Classifier, Safe drugs.

Abstract
The Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Random Forest is a novel approach aimed at assisting healthcare professionals in making informed decisions regarding drug prescriptions. This system leverages the power of sentiment analysis to analyze and extract valuable insights from large volumes of user-generated drug reviews. Sentiment analysis plays a crucial role in understanding the opinions, emotions, and experiences expressed by patients in their reviews. By classifying these sentiments as positive, negative, or neutral, the system provides a quantitative measure of the overall sentiment associated with a particular drug. This information can then be utilized to generate personalized drug recommendations tailored to individual patients' needs and preferences. The proposed system employs the Random Forest algorithm, a powerful machine learning technique, to perform sentiment analysis on drug reviews. Random Forest utilizes an ensemble of decision trees to classify sentiments based on a variety of features extracted from the reviews, such as keyword frequencies, sentence structures, and context. By aggregating the results of multiple decision trees, Random Forest achieves high accuracy and robustness in sentiment classification. To build the Drug Recommendation System, a comprehensive dataset of drug reviews is collected from various online sources. The dataset is pre-processed to remove noise, perform text normalization, and extract relevant features. Then, the Random Forest algorithm is trained on this processed dataset, using labelled reviews as input and their corresponding sentiments as output. The trained model is subsequently used to analyse new, unseen drug reviews and predict their sentiments. Once the sentiment analysis is performed, the system combines the sentiment scores with additional factors, such as drug efficacy, safety, side effects, and patient demographics, to generate personalized drug recommendations. These recommendations are based on a similarity measure that matches patients with similar profiles and preferences to those who have reported positive experiences with certain drugs.

1. Introduction

A drug recommendation system is an application of artificial intelligence (AI) in healthcare that assists healthcare providers in making personalized drug treatment recommendations for their patients. The system uses machine learning algorithms to analyse patient data, such as electronic health records (EHRs), medical history, lab results, and genetic information, to provide treatment recommendations that are specific to each patient's needs. The system can help healthcare providers choose the most appropriate medication for a patient based on their medical history, symptoms, and other relevant factors. The goal of the drug recommendation system is to improve patient outcomes by reducing adverse drug events, minimizing drug interactions, and ensuring patients receive the most effective treatment. The drug recommendation system can use a variety of machine learning algorithms, including supervised learning, unsupervised learning, deep learning, and natural language processing (NLP). These algorithms can analyse patient data and provide treatment recommendations based on patterns and trends in the data. The drug recommendation system can be integrated into existing healthcare software, such as electronic health record (EHR) systems, to provide healthcare providers with seamless access to treatment recommendations. The system can also be integrated with clinical decision support systems (CDSS), which provide clinical guidelines and treatment recommendations to healthcare providers. The use of a drug recommendation system can help reduce healthcare costs by minimizing unnecessary testing and

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

procedures and reducing the length of hospital stays. The system can also improve patient satisfaction by providing personalized treatment recommendations that are specific to each patient's needs. Overall, the drug recommendation system is a promising application of AI in healthcare that has the potential to improve patient outcomes, reduce healthcare costs, and enhance the quality of care.

2. Literature Survey

Witch CM et.al [9] With an emphasis on nomenclature, definitions, incidence, risk factors, disclosure, and legal ramifications, this article reviews pharmaceutical errors for general practitioners. Medication errors can be caused by a variety of factors, such as those relating to the drug, the patient, and the healthcare professional. Losing the faith of their patients, legal litigation, criminal accusations, and medical board reprimand are just a few of the consequences doctors may experience after making prescription mistakes. To prevent pharmaceutical errors, a variety of strategies have been tested, with various degrees of effectiveness. By knowing more about pharmaceutical errors, doctors may better ensure that their patients receive safe care. J.G. Bartlett et al. An Intelligent Medicine Recommender System Framework [2]-In this paper, the author introduces a universal medicine recommender implemented to apply data mining technologies to recommendation systems. This system improves improve patient health and enhance safety by providing tailored drug recommendations that take into account individual patient characteristics, such as age, gender, weight, and medical history. This system may not be able to cover all possible drug interactions, contraindications, and other safety concerns. This recommendation model does not provide most up-to-date drugs. Medicine Recommendation System Based on Patient Reviews [8]-This pharmaceutical recommendation algorithm, which is based on patient reviews, suggests all medications based on their average anticipated value. When the mean projected value is higher, the medicine is more accurate. Therefore, by taking into account the greatest mean value predicted, doctors prescribe the best medication. The age of the person or other demographic data is not taken into account while making recommendations using this technique. Additionally, the brand and chemical components of the drug need to be improved.

3. Existing Work

The existing system comprises of a machine learning based drug recommendation system using Content based filtering technique which involves using features of the drugs themselves to recommend other drugs with similar properties. It was developed using a sample dataset from Kaggle over Google Colab platform. In this work, the development of a Drug Recommender System (DRS) for different diseases to maintain good patient health and longevity. They addressed the unfairness in drug usage by DRS for severe chronic diseases by improving the recommendation accuracy by the integration of ML knowledge [1]. The goal of this recommendation system is to examine the dataset using data mining concepts, visualization, sentiment analysis and recommend drugs based on the condition, ratings and reviews using Machine Learning approaches, Content filtering approach, for each health condition of a patient [3]. The pre-processed data is trained and tested using linear Svc classifier. Finally, the trained model is validated and analysed with test data and provide predicted results.

In the system, the drug is offered on a specific condition dependent on patient reviews and ratings using technologies like Machine Learning classifier such as Logistic Regression (0.89), Perceptron (0.898), Ridge classifier (0.892), Multinomial NB (0.881), SGD Classifier (0.878), Linear Svc (0.90). The maximum accuracy achieved in the existing system is Linear Svc. This Linear SVC classifier answers to the question whether the key feature present or not and which takes long time to train the model. Drug recommendation systems rely on large amounts of data in order to make accurate recommendations. However, obtaining large amounts of high-quality data can be challenging, particularly in areas such as drug development where data may be limited. The result shows that this framework improves the cure rate. In this research, multilingual sentiment analysis was performed using Linear SVC and Recurrent Neural Network (RNN). This model suffers from the disadvantages of vanishing gradients, long term memory and difficulty in parallelization.

4. Proposed Method

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

The proposed drug recommendation system is an established method that, based on sentimental analysis and the user's health situation, suggests particular medications We began by gathering data for the drug review dataset from Kaggle and the UCI ML repository. Six attributes are included in this dataset. Preparing the data for analysis requires numerous processes in the pre-processing of the data includes handling missing values, Summarising the dataset, Revealing hidden patterns from the data and engineering of features. The useful count and useless count are segmented based on the number of users who found the drug review helpful or not. The useful count represents the number of users who found the drug review helpful, while the useless count represents the number of users who did not find the drug review helpful. This model involves training a LSTM model to classify drug reviews based on their sentiment (negative, neutral, positive). After that, it pads sequences and converts rating labels to three classes (negative, neutral, positive). The sentiment analyser used is the Vader Lexicon from NLTK. It returns a dictionary containing the positive, negative, neutral, and compound scores for the given text. The resulting value represents the usefulness of the drug, with higher values indicating that the drug is more useful. The predicted sentiment was measured using five metrics, namely, precision (Prec), recall (Rec), f1 score (F1), accuracy (Acc.) After assessing the metrics, all four best-predicted results were picked and joined together to produce the combined prediction and recommend top five drugs based on health condition by users.

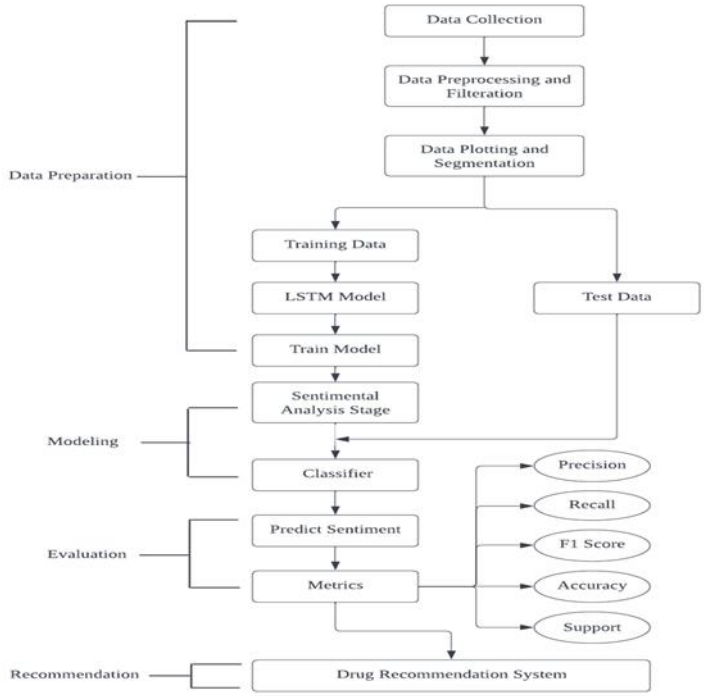

Figure.1 Architectural Diagram for Proposed Work

A. Data Collection & Data Pre-processing

The drug review dataset, which was obtained from the UCI ML repository and Kaggle, has six attributes: drugName, which gives the name of the drug; condition, which gives the name of the condition related to that drug; review, which provides reviews from patients who have used the drug for that particular condition; rating, which is a score out of 10 given by the patients to the drugs; date, which provides the date when the review was posted; and usefulCount, which indicates the number of reviews that were actually useful.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Attribute Name	Type	Description
drugName	Categorical	Name of the drug
Condition	Categorical	Name of condition
Review	Text	Patient review
Rating	Numerical	10 star patient rating
Date	Date	Date of review entry
usefulCount	Numerical	Number of users who found review useful

Table 1. Attributes of dataset

Preparing the data for analysis requires numerous processes in the pre-processing of the data. The steps are as follows: Reading the data: Using the pandas library, read the data from a CSV file in the first step. Handling missing values: Handling missing values in the dataset is the next stage. Summarising the dataset: To obtain statistical data on the numerical columns in the dataset, the dataset is then summarised. Revealing hidden patterns from the data: Using a variety of visualisation techniques, including scatter plots, bar plots, and histograms, the data is then examined to reveal hidden patterns. Feature engineering is performed to create new features from the existing ones.

B. Data segmentation and plotting

The useful count and useless count are segmented based on the number of users who found the drug review helpful or not. The useful count represents the number of users who found the drug review helpful, while the useless count represents the number of users who did not find the drug review helpful. The code uses the 'matplotlib' and 'seaborn' libraries to create various plots to analyze the data. Some of the plots created are: - Distribution of Rating and Useful Count: This plot shows the distribution of the 'rating' and 'usefulCount' columns using a histogram. Rating vs Usefulness: This plot shows the impact of ratings on usefulness. Length of Review vs Ratings: This plot shows whether the length of the review has any impact on the ratings of the drugs. It creates a new column named 'len' to calculate the length of the reviews create a scatter plot with 'len' on the x-axis and 'rating' on the y-axis. As shown in fig. 3 we have found the top 10 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.

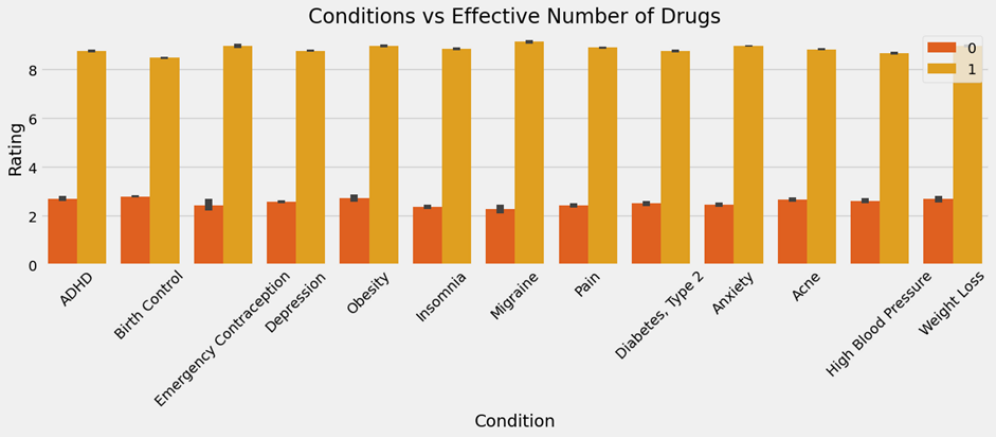

Figure 2. Conditions vs Effective No. of Drugs

C. Training Reviews using LSTM

This model involves training a LSTM model to classify drug reviews based on their sentiment (negative, neutral, positive). It uses Tokenizer to assign token values to words in the user reviews. After that, it pads sequences and converts rating labels to three classes (negative, neutral, positive).The neural network model

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

is defined with Sequential() and various layers such as Embedding, Conv1D, MaxPooling1D, LSTM, Dropout, and Dense are added. It is compiled using categorical_cross entropy as the loss function, Adam as the optimizer, and accuracy as the metric. Finally, the model is trained on train_data and train_labels using a batch size of 128 and validation split of 0.2.

```
Epoch 1/5
1009/1009 [==============================] - 183s 176ms/step - loss: 0.6274 - accuracy: 0.7671 - val_loss: 0.5425 - val_accuracy: 0.7976
Epoch 2/5
1009/1009 [==============================] - 162s 161ms/step - loss: 0.5083 - accuracy: 0.8149 - val_loss: 0.5251 - val_accuracy: 0.7998
Epoch 3/5
1009/1009 [==============================] - 164s 163ms/step - loss: 0.4595 - accuracy: 0.8306 - val_loss: 0.4926 - val_accuracy: 0.8189
Epoch 4/5
1009/1009 [==============================] - 169s 167ms/step - loss: 0.4212 - accuracy: 0.8432 - val_loss: 0.4794 - val_accuracy: 0.8218
Epoch 5/5
1009/1009 [==============================] - 169s 168ms/step - loss: 0.3940 - accuracy: 0.8526 - val_loss: 0.4807 - val_accuracy: 0.8260
```

Figure 3. Model trained for 5 epochs

D. Sentimental Analyzer Stage

The sentiment analyzer used is the Vader Lexicon from NLTK. It is implemented using the SentimentIntensityAnalyzer class from the nltk.sentiment.vader module. The analyzer calculates the sentiment of each review in the dataset using the polarity scores() method, which returns a dictionary containing the positive, negative, neutral, and compound scores for the given text. The compound score is a normalized score that ranges from -1 (most negative) to +1 (most positive), with 0 being neutral. The compound score is used as the sentiment score for each review in the dataset. the effectiveness of drugs is calculated using the `scale_rating` function, which takes the `rating` of a drug as input and returns a binary value of 0 or 1 based on whether the drug is considered effective or not. The function multiplies the `rating` by 5, rounds it to the nearest integer, and then checks if the rounded value is 0, 1, or 2. If it is, the function returns 0, indicating that the drug is not effective. Otherwise, it returns 1, indicating that the drug is effective. The usefulness of drugs is calculated by multiplying the `rating`, `usefulCount`, and `eff_score` columns of the dataset together. This calculation takes into account the rating of the drug, the number of times it has been marked as useful by users, and its effectiveness score. The resulting value represents the usefulness of the drug, with higher values indicating that the drug is more useful.

E. Model Evaluation and Drug Recommendations

The predicted sentiment was measured using five metrics, namely, precision (Prec), recall (Rec), f1 score (F1), accuracy (Acc.) After assessing the metrics, all four best-predicted results were picked and joined together to produce the combined prediction. The most useful and useless drugs for each condition can be found using drug recommendation system. This takes a list of conditions as input (by default, it takes all the conditions in the dataset) and prints the top 5 drugs for each condition based on their usefulness score. The usefulness score is calculated as the product of the drug's rating, useful count, and effectiveness score. The function sorts the drugs in descending order of their usefulness score to find the most useful drugs and in ascending order to find the least useful drugs.

5. Results

Each review in this work was categorised as either positive or negative based on the user's star rating. Positive reviews are defined as those with scores of five or higher, while a scale of one to five stars. Initial training data contained 161298 ratings and 33000 testing data contained ratings, respectively. We have analysed different algorithms on the dataset to understand which one gives the better accuracy. The algorithms used are Naive Bayes, Linear SVC, Logistic Regression and LSTM. The algorithm giving the best accuracy was selected to train the dataset. The accuracy of the algorithm are given in the table 2. The best accuracy is given in bold.

Algorithm	Accuracy
Multinomial Naive Bayes	0.75354
Linear SVC	0.58199

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Logistic Regression	0.63778
LSTM	**0.82601**

Table 2. Comparison of different algorithms

The long short term memory algorithm processes the entire sequence of data as they have gates which regulates the flow of information.

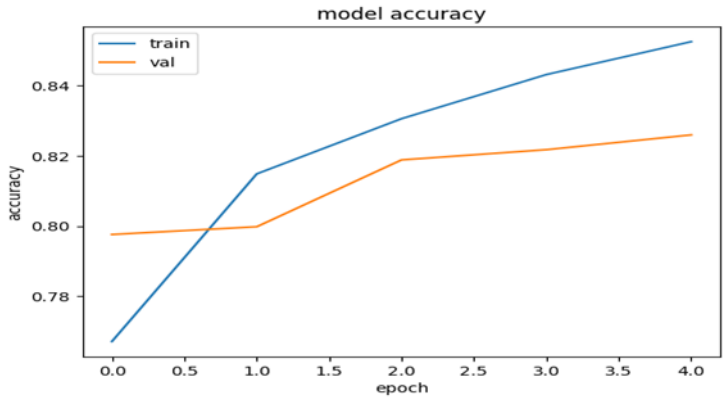

Figure 4. Model accuracy visualization

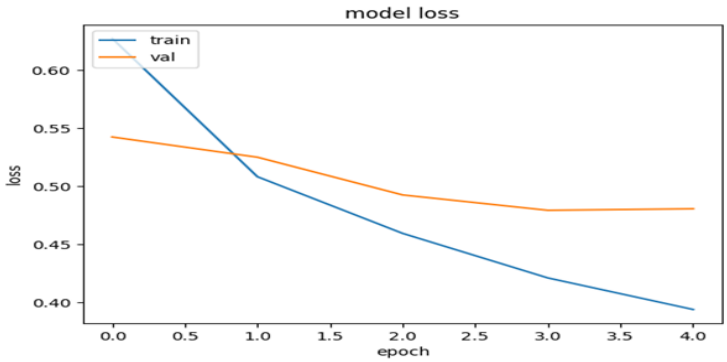

Figure 5. Model loss visualization

In this drug recommendation system based on sentimental analysis 70 % percentage of the data set are used for training and 30% dataset are used for testing. This model is trained for 5 epochs with the batch size of 128 whose model accuracy is 85% and model loss is 39%. In this project, the proposed work of sentimental analyzer module is implemented using the LSTM algorithm with the precision, Recall, F1-Score, Support.

Sentiment	Precision	Recall	F1-Score	Support
Negative	77%	78%	78%	13497
Neutral	39%	10%	17%	4829
Positive	86%	94%	90%	35440

Table 3. Result Analysis

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

The final output shows the top five drugs recommended by our model on the condition, Birth control.

```
Birth Control

 Top 5 Drugs
                                drugName  usefulness
0                           Tri-Sprintec        1190
1  Ethinyl estradiol / norelgestromin         1150
2                           Etonogestrel         936
3   Ethinyl estradiol / norethindrone           900
4                               TriNessa         810
```

Figure 6. Drug Recommendation System

6. Conclusion

In conclusion, the Drug Recommendation System based on Sentiment Analysis of Drug Reviews is a promising approach to assist healthcare professionals, patients, and consumers in making informed decisions about medication usage. By leveraging machine learning and natural language processing techniques, the system analyzes user-generated drug reviews to extract sentiments and provide personalized recommendations.

The advantages of this system are significant. It offers data-driven recommendations based on a large volume of real-world experiences, leading to evidence-based decision-making. The personalized nature of the recommendations ensures that individuals find drugs that align with their specific needs and preferences. Moreover, the system improves the efficiency of decision-making, saving time and costs associated with manual analysis of reviews.

One of the notable strengths of this system is its ability to enhance patient safety. By analyzing sentiments and identifying potential adverse reactions or side effects, healthcare professionals and patients can make informed choices, minimizing risks and improving overall treatment outcomes.

However, there are certain limitations to consider. Biases present in the collected data can affect the accuracy and representativeness of recommendations. Sentiment analysis techniques may struggle with understanding nuanced sentiments, sarcasm, or complex language structures. Random Forest models may overgeneralize patterns, leading to potential misclassifications.

To address these limitations, it is crucial to continuously improve data collection processes, ensuring diverse and representative datasets. Advancements in natural language processing can help in better understanding contextual nuances and improving sentiment analysis accuracy. Ongoing research and development efforts are needed to refine the algorithms and address bias and overgeneralization issues.

It is important to note that the Drug Recommendation System should not replace professional medical advice. The system should be used as a supportive tool, with users consulting healthcare professionals for comprehensive guidance.

In summary, the Drug Recommendation System based on Sentiment Analysis of Drug Reviews great potential to enhance decision-making in medication usage. By harnessing the power of machine learning and sentiment analysis, this system provides personalized, data-driven recommendations that can improve patient care, promote informed prescribing decisions, and empower individuals to make educated choices about their medications. Continuous advancements in technology and addressing limitations will further enhance the effectiveness and reliability of this system in the future.

References

[1] Y. Bao and X. Jiang, “An intelligent medicine recommender system framework,” in Proceedings of the 2016 IEEE 11th Conference on Industrial

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Electronics and Applications, ICIEA 2016, Oct. 2016, pp. 1383–1388, doi: 10.1109/ICIEA.2016.7603801

[2] 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.Y

[3] D. Anil, A. Vembar, S. Hiriyannaiah, S. G.M. and K. G. Srinivasa, "Performance Analysis of Deep Learning Architectures for Recommendation Systems," 2018 IEEE 25th International Conference on High Performance Computing Workshops (HiPCW), Bengaluru, India, 2018, pp. 129-136.

[4] R. C. Chen, Y. H. Huang, C. T. Bau, and S. M. Chen, "A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection," Expert Syst. Appl., vol. 39, no. 4, pp. 3995–4006, Mar. 2012, doi: 10.1016/j.eswa.2011.09.061.

[5] J.-C. Na and W. Y. M. Kyaing, "Sentiment Analysis of User-Generated Content on Drug Review Websites," J. Inf. Sci. Theory Pract., vol. 3, no. 1, pp. 6–23, Mar. 2015, doi: 10.1633/jistap.2015.3.1.1.

[6] Soni J, Ansari U, Sharma D, et al. Predictive data mining for medical diagnosis: An overview of heart disease prediction[J]. International Journal of Computer Applications, 2011, 17(8): 43-48.

[7] 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.

[8] T. Venkat Narayana Rao, Anjum Unnisa, Kotha Sren International Journal Of Scientific & Technology Research Volume 9, Issue 02, February 2020 Issn 2277-8616

[9] Doulaverakis, C., Nikolaidis, G., Kleontas, A. et al. GalenOWL: Ontology-based drug recommendations discovery. J Biomed Sema

[10] Wittich CM, Burkle CM, Lanier WL. Medication errors: an overview for clinicians. Mayo Clin Proc. 2014 Aug;89(8):1116-25.