Drug Recommendation System Using Machine Learning

[1]Asadi Muni Hemanth, [2]Dr. R. Swathi, [3]Dr. K. Santhi

[1]M. Tech Student, [2,3]Professor
Department of Computer Science
Sri Venkateswara College of Engineering, Tirupati

***Abstract-* Since corona virus has shown up, inaccessibility of legitimate clinical resources isat its peak, like the shortage of specialists and healthcare workers, lack of proper equipment and medicines etc. The entire medical fraternityis in distress, which results in numerous individual's demise. Due to unavailability, individuals started taking medication independently Without appropriateconsultation, 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 project intends to present a drug recommendation system based on sentiment analysis that can drastically reduce specialist's heap. In this project, we built a drugrecommendation system that uses patient reviews to predict the sentiment which lets people know whether to take a particular drug or not.**

***Keywords*: corona virus, sentiment analysis , machine learning**

I. INTRODUCTION

With the number of coronavirus cases growingly exponentially, the nations are facinga shortage of doctors, particularly in rural areaswhere 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 Telemedicine framework ought to be energized as far as possible in this difficult time.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. Over40% medicine, specialists make mistakes whileprescribing since specialists compose the solution as referenced by their knowledge, which is very restricted. Choosing the top- based information about microscopic organisms, antibacterial medications, andpatients.

Every day a new study comes up with accompanying more drugs, tests, accessible forclinical staff every day. Accordingly, it turns outto be progressively challenging for doctors to choose which treatment or medications to give to a patient based on indications, 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. Common causes of medication error incorrect diagnosis. Prescribing errors, dose miscalculations, poor drug distribution practices, drug and drug device related problems, incorrect drug administration, failed communication and lack of patient education. Medication errors are among the most common medical errors, harming at least 1.5 millionpeople every year. The extra medical costs of treating drug-related injuries occurring in hospitals alone are at least to $3.5 billion a year,and this estimate does not take into account lostwages and productivity or additional health carecosts, the report says. Medication error morbidity and mortality costs are estimated to run $77 billion dollars per year. Patient safety ismajor public health concern.

The Academy of Managed Care Pharmacy (ACMP) recognized the importance of this issue and supports programs that help achieve the goal of improved patient safety andprevention of medication errors. ACMP'sFramework for Quality Drug Therapy, emphasizes and promotes public safety, continuous monitoring for accuracy in dispensing, reliability in the transmission of prescription and medication orders, and continuous review and upgrade of pharmacy operating system.

Individuals worldwide become adjusted to analyze reviews and websites first before settling on a choice to buy a thing. Therehas been an expansion in the number of individuals in a Pew American Research centersurvey directed in 2013, roughly 60% of grown-ups searched online for health-related subjects,and around 35% of users looked for diagnosinghealth conditions on the web. A medication recommender framework is truly vital with thegoal that it can assist specialists and help patients to build their knowledge of drug on specific health conditions.

A recommender framework is acustomary system that proposes an item to the user, depends on their advantage and necessity.The data which is being used in this study is analyzed in two main ways; as categorical dataand as numerical data. The dataset originally comes with categorical data. The raw data can be prepared by data cleaning and other basic preprocessing techniques. First, categoricaldata can be transformed into numerical data andthen appropriate techniques are applied to do the evaluation. Secondly, categorical data isused in the machine learning techniques to findthe optimal algorithm. 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 usingsentiment analysis.

Therefore, a medication recommender framework is truly vital with the goal that it canassist specialists and help patients to build theirknowledge of drugs on specific health conditions.

II. EXISTING SYSTEM

In the existing system, implementation ofmachine learning algorithms is bit complex tobuild due to the lack of information about thedata visualization. Mathematical calculationsare used in existing system for model buildingthis may takes the lot of time and complexity. The Existing System that is used to predict thedrug review performs the following steps topredict the sentiment:

1. Collection of the data from varioussources.
2. Classification of the data under suitableheadings.
3. Analyzing the data.
4. Predict the output.

Drawbacks of Existing System

- No proper user interface to useregularly.
- Prediction of results are not soaccurate.
- Very few systems use the availablereview data for the prediction purposes.
- Time consuming.
- Difficulty in implementation, implementing the existing machine learning model can be challenging. This can be due to the factors like unfamiliar technology.
- Lack of technical knowledge amongmost of the individuals to use the prediction system.

III. PROPOSED SYSTEM

Proposed several machine learning models toclassify whether the medicine has positive sentiment or negative sentiment, but none haveadequately addressed this misdiagnosisproblem. Also, similar studies that have proposed models for evaluation of such performance classification mostly do not consider the heterogeneity and the size of the data. To overcome all this, we use machine learning packages available in the scikit-learn library. Therefore, we propose a Decision Tree,Logistic Regression to classify the performance.

The advantage of Machine Learning in Drug Review Classifier is that it keeps on improvingas it is exposed to more data.

The proposed algorithms used in this project is:

- Logistic Regression
- Decision Tree

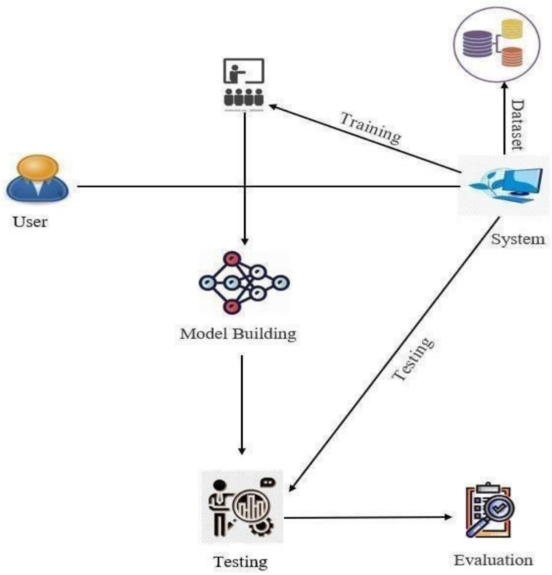
Features of Proposed System

- Well-developed user interface to use regularly**.**
- Prediction of results are more accurate.
- Uses Logistic Regression for more accurate results.
- Less Time consuming.

Advantages of Proposed System

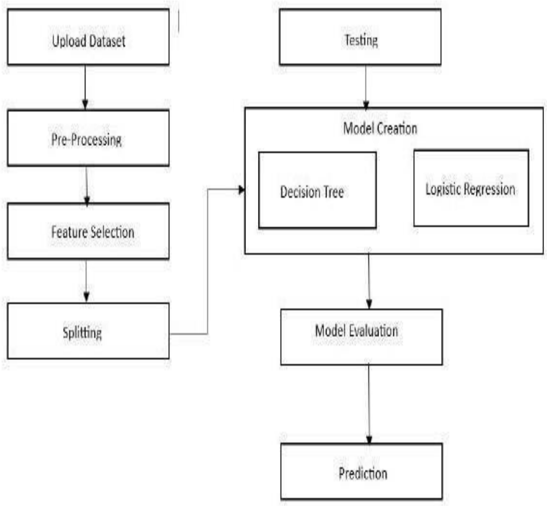
- No Medical and technical knowledge isrequired.
- Prediction accuracy is increased.
- Reduce the time complexity of doctors.

IV. SYSTEM ARCHITECTURE



V. BLOCK DIAGRAM

- First the user needs to upload the dataset.
- Then the data is converted in to binary dataset.
- Pre-processing is performed on data. Itsplits the data into two parts.
- By using ML algorithms, training of thedata is performed.
- Finally, after training modelling it candetect the sentiment of drug.



VI. IMPLEMENTATIONModules

- Data Collection and Preprocessing
- Splitting Data
- Training the model
- Model Evaluation
  - Prediction MODULES DESCRIPTION:

Data Collection and Preprocessing

The drug Review Data is collected from the online sources such as Kaggle and other sites inthe csv format. The Drug Review Data consistsof several parameters such as Drug unique ID, Drug name, condition, Review, Rating.Collected data is processed to remove any missing values, noises, duplicates in the data set. Processing of data is required for improvingthe accuracy of the model.

Splitting Data

After the collection of the dataset, we split the dataset into Training Data and Testing Data. The Training Dataset is used for training the prediction model and Testing Dataset is used forevaluating the prediction model. For thisproject the Data set is split as 80% and 20%. The 80% is used for training the prediction model and 20% is used for testing the predictionmodel.

Training the Model

For Training the prediction model 80% of the collected dataset is used. Logistic Regression Machine Learning Algorithm is used to train theprediction model which is more accurate. It is used for predicting the categorical. Logisticregression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can beeither Yes or No, 0 or 1, true or False, etc. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).

Model Evaluation

For Evaluating the prediction model 20% of thecollected dataset is used. The metric used for evaluating the model in the project is accuracy score. Accuracy score is the percentage of correctly predict instances in the data set. It is calculated by dividing the number of correctly predicted instances by the total number of instances in the data set.

Prediction

The prediction System is built using the Trainedand Evaluated model. Building process of the Heart Disease Prediction System includes all the above steps such as Data Collection and Processing, splitting of the data, Training the model, Model Evaluation. The Combination ofall these processes forms the Prediction System.

The module contains several parameters as input and predicts the sentiment of review.

VII. ALGORITHMSDecision Tree:

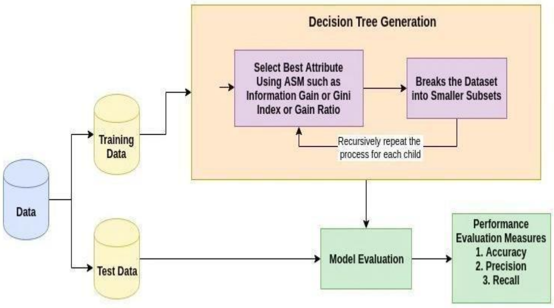
Decision tree is a flowchart-like tree structure where an internal node represents feature (or attribute), the branch represents a decision rule,and each leaf node represents the outcome. Thetopmost node in a decision tree is known as theroot node.

The basic idea behind any decision treealgorithm is as follows:

1. Select the best attribute using AttributeSelection Measures (ASM) to split the records.

2. Make that attribute a decision node andbreaks the dataset into smaller subsets.
3. Starts tree building by repeating this process recursively for each child untilone of the conditions will match:

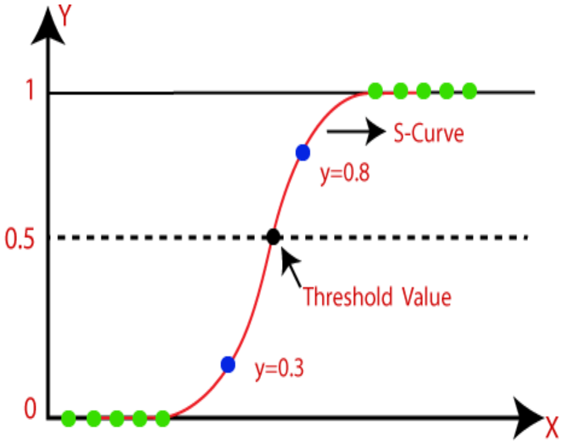
- All the tuples belong to thesame attribute value.
- There are no more remainingattributes.
- There are no more instances.



Logistic Regression:

Logistic Regression is used when the dependentvariable (target) is categorical. Logisticregression is one of the most commonly used machine learning algorithms for binary classification problems, which are problems with two class values, including predictions such as “this or that,” “yes or no” and “A or B”.

The purpose of logistic regression is to estimatethe probabilities of events, including determining a relationship between features and the probabilities of particular outcomes.



Logistic Function

The Logistic regression equation can be obtained from the Linear Regression equation. The mathematical steps to get Logistic Regression Equations are given below:

- We know the equation of the straight line can be written as:
- ❖ Y=b0+b1x1=b2x2+b3x3+…+bnxn
- In Logistic Regression y can be between 0 and 1 only, so for this let’s divide the above equation by (1-y):
- ❖ y/1-y; o for y=0, and infinity for y=1
- But we need range between –[infinity]to + [infinity], then take logarithm of the equation it will become:
- ❖ Log[y/1-y]= b0+b1x1+b2x2+b3x3+…+bnxn

The above equation is the final equation for Logistic Regression.

VIII. 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 studiedto build a recommender system using different types of machine learning classifiers, such as Logistic Regression and Decision Tree.

Designing and Developing a DrugRecommendation System Using Machine Learning algorithm to get more accurate resultsand raise awareness among the people and helpindividuals to overcome.

The overall aim is to create a Drug Recommendation System that suggests the proper drug to patients. In this system we implement logistic regression to predict the outcome accurately and also provide a Graphical User Interface so that the user can usethe system with ease whenever and where everthey want.

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. 2014Aug;89(8):1116-25.
3. CHEN, M. R., & WANG, H. F. (2013). Thereason and prevention of hospital medication errors. Practical Journal of Clinical Medicine, 4.
4. Drug Review Dataset, https://archive.ics.uci.edu/ml/datasets/Dru g% 2BReview%2BDataset%2B%2528Drugs.com%2529#
5. Fox, Susannah, and Maeve Duggan.“Health online 2013. 2013.” URL: http://pewinternet.org/Reports/2013/Healt h-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.
7. Clin Infect Dis. 2000 Aug;31(2):347-82. Doi: 10.1086/313954. Epub 2000 sep 7. PMID: 10987697; PMCID: PMC7109923.
8. Fox, Susannah & Duggan, Maeve. (2012). Health Online 2013. Pew Research InternetProject Report.
9. 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.
10. Doulaverakis, C., Nikolaidis, G., Kleontas,
11. A. et al. GalenOWL: Ontology-based drugrecommendations discovery. J Biomed Semant 3, 14 (2012). https://doi.org/10.1186/2014-1480-3-14
12. Leilei Sun, Chuanren Liu, Chonghui Guo, Hui Xiong, and Yanming Xie. 2016. Data-driven Automatic Treatment Regimen Development and Recommendation. In Processings 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.
13. V. Goel, A. 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.
14. Shimada K, Takada H, Mitsuyama S, etal. Drug-recommendation system for patients with infectious diseases. AMIA Annu Symp Proc. 2005;2005:1112.
15. Y. Bao and X. Jiang, “An intelligent medicine recommender systemframework,” 2016 IEEE 11th Conference on Industrial Electronics and Applications(ICIEA), Hefei, 2016, pp. 1383-1388, doi: 10.1109/ICIEA.2016.7603801.
16. Zhang, Yin & Zhang, Dafang & Hassan, Mohammad & Alamri, Atif & Peng, Limei. (2014). CADRE: Cloud- Assisted Drug Recommendation Service for Online Pharmacies. Mobile Networks and Applications. 20.348-355.
17. 10.1007/s11036-014-0537-4.
18. J. Li, H. Xu, X. He, J. Deng and X. Sun,“Tweet modeling with LSTM recurrent neural networks for hashtagrecommendation,” 2016 International JointConference on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 1570-1577, doi:10.1109/IJCNN), Vancouver, BC, 2016, pp.1570-1577, doi:
19. 10.1109/IJCNN.2016.7727385.
20. Zhang, Yin & Jin, Rong & Zhou, Zhi- Hua. (2010). Understanding bag-of-words model: A statistical framework.International Journal of Machine Learning and Cybernetics. 1. 43-52.
21. 10.1007/s13042-010-0001-0.
22. J. Ramos et al., “Using tf-idf todetermine word relevance in document queries,” in Proceedings of the first instructional conference on machine learning, vol. 242, pp. 133-142, Piscataway,NJ, 2003.
23. Yoav Goldberg and Omer Levy. Word2vec Explained: deriving Mikolov et al.’s negative-sampling word-embedding method, 2014; arXiv:1402.3722.
24. Danushka Bollegala, Takanori Maehara and Kenichi Kawarabayashi. Unsupervised Cross-Domain WordRepresentation Learning, 2015;arXiv:1505.07184.
25. Textblob, https://textblob.readthedocs.io/en/dev/.